Cellular mRNA Export Factor UAP56 Recognizes Nucleic Acid Binding Site of Influenza Virus NP Protein

By

Andrew Kennedy Morris

Dissertation

Submitted to the Faculty of the

Graduate School of Vanderbilt University

in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in Chemical and Physical Biology

May 31, 2020

Nashville, TN

Approved:

Dana Borden Lacy

Charles Sanders

Erkan Karakas

Lauren Parker Jackson

Manuel Ascano

TABLE OF CONTENTS

Page

LIST OF FIGURES

Figure Page

CHAPTER 1

INTRODUCTION

1.1 – Introduction

One of the defining features of eukaryotes is the separation of the transcription and processing of genetic material from its translation by the physical separation of the former into the nucleus and the latter in the cytoplasm. This provides for a much greater degree of regulation and control over the transcription process than that seen in prokaryotes. This also means that mRNAs must mature and be exported, and this involves an intricate set of cellular machinery to coordinate fine control over transcription initiation, 5' end capping, splicing, cleavage and polyadenylation, and translocation through the nuclear pore complex. The Ren lab is principally focused on understanding how the pieces of this machinery fit together at an atomic and structural level, and the functional implications of these interactions on mRNA processing and export. The lab particularly has an interest in the mis-regulation of the processing and export machinery and its role in disease. Often the more detailed functions of a biological process are discovered when observed as a phenotype resulting from the disruption of it. In the field of mRNA export, this has meant that several notable discoveries, including that of the primary bulk mRNA export factor NXF1/NXT1, were made based on observed interaction with viruses[1,2].

Most viruses replicate in the cytoplasm, however the influenza virus is among those that replicate in the nucleus. One advantage this presents is the evasion of most of the host innate immune factors during transcription and replication, as these factors are generally confined to the

cytoplasm. This however also means that they have had to evolve ways to traffic components into and out the nucleus, including their mRNA transcripts. Because of this, study of influenza-host interactions can provide insight into the more fundamental aspects of mRNA processing and export.

This project looks at the interaction of a critical mRNA export factor UAP56, and an influenza viral protein called nucleoprotein. In examining the mechanism by which a virus exploits the host transcriptional process, we add to both the fields of influenza biology and mRNA processing and export. The relevant backgrounds of these fields, and their intersection is presented in chapter 1, while chapter 2 discusses our efforts to examine the minimal binding regions and the conditions which affect the UAP56 NP interaction. Chapter 3 discusses the methods to obtain an atomic resolution view of the interface between this binding pair and the role of RNA in this interaction. The final chapter concludes with a summary of findings, as well as the open questions which remain. The majority of the work described here was originally published by us in Morris *et al,* 2020[3], and figures from this paper are noted in their figure legends.

1.2 – Host mRNA Processing and Export

1.2.1 – Overview

This project is at the interface of two different fields, the basic level biochemistry of influenza and the field of mRNA processing. These fields are highly intertwined, since influenza, or indeed any virus, is an obligate parasite, and a complete understanding of influenza replication is contingent on a knowledge of host mRNA processing and export. The aims of the lab are

fundamentally to understand processes in mRNA export, with the study of viruses being secondary, and so it is this area which this dissertation will begin with by reviewing the basic processes involved, beginning at transcription, through to the export of mature mRNA through the nuclear pore.

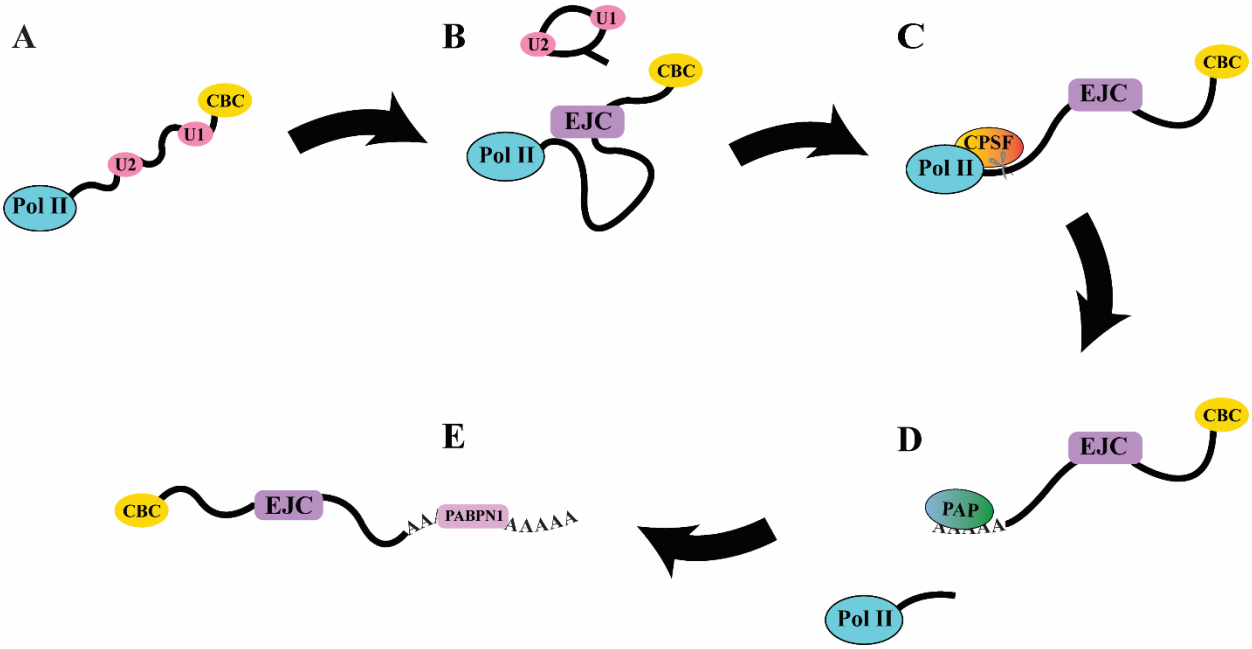

Figure 1.1 – mRNA Processing

A) In early transcription, the 7-methyl guanosine cap is bound by the Cap Binding Complex and U1 and U2 snRNPs bind to exonic junctions. **B)** Splicing occurs and its associated factors disengage, leaving the EJC. **C)** When transcription reaches the polyadenylation sequence AAUAAA, Cleavage and Polyadenylation Specificity Factor (CPSF) cleaves the transcript from Pol II elongation **D)** The free transcript dissociates from Pol II, which eventually terminates transcription. Poly-A polymerase (PAP) adds a variable number of adenines to the 3' end of the nascent transcript. **E)** Poly-A Binding Protein Nuclear 1 (PABPN1) binds to the new 3' poly-A tail.

1.2.2 – Capping

mRNA processing begins after transcription initiation. As soon as mRNA emerges from pol II, it is almost immediately subject to 5' capping. A defining feature of eukaryotic mRNAs is the addition of the unusual nucleotide 7-methyl guanosine (7mG) triphosphate in a 5' to 5' linkage to the first base of a transcript. This capping at the 5' end is essential to processing and allows the transcript to avoid rapid exonuclease degradation[4]. In metazoans, this is carried out by the enzyme RNGTT[5]. The methylation of the terminal guanine is catalyzed by RNA (guanine-7-)methyltransferase and RNMT-activating mini-protein in vertebrates[6]. 7mG is then bound by a large multiprotein complex called the Cap Binding Complex (CBC) which plays a role in recruiting other processing factors (see **Fig. 1.1 A**). It also interacts with a complex network of transcription factors which form a feedback loop to pol II which can enhance or sometimes repress premature transcription termination depending on the transcript[5].

As the transcript elongates, heterologous nuclear ribonucleoproteins (hnRNP) bind along it to stabilize it. hnRNPs are a large class of small RNA binding proteins, each with varying binding sequences and specificities. These coat much of the sequence of an mRNP and the specific set of hnRNPs bound to a transcript is known to play a large part in regulating splicing, export, and later in the cytoplasm, translation[7].

1.2.3 – Spliceosome Recruitment

The vast majority of transcripts also undergo splicing, that is, the removal of introns and variable incorporation of exons. Incorporation of different exons through alternative splicing is one method of controlling gene expression and function through the production of different isoforms,

while introns produced are the source of snoRNAs[8] and often have regulatory roles as sources of lncRNAs and miRNAs[9]. Splicing proceeds through a very large and complex structure called the spliceosome which contains a core of large RNA-protein complexes called snRNPs with many different auxiliary factors which are associated with different steps in catalysis[10]. An oversimplified mechanism of splicing is as follows. A 5' exonic boundary consisting of a guanine followed by a uridine is bound by the U1 snRNP, while an internal branch point adenine is bound by the U2 snRNP. The U4, U5, and U6 snRNPs form a bridge between these, and an ATP-dependent rearrangement occurs that brings the U1 and U6 sites, now holding the 5' exonic boundary, together. A transesterification reaction occurs whereby the 2' hydroxyl of the guanine at the 5' exonic boundary attacks the internal branch point adenine, leading to a loop structure with the adenine sugar engaged in three different phosphodiester bonds. This creates a new free 3' end at the 5' exonic boundary, which engages in a second transesterification with another guanine at the 3' exonic boundary downstream. This joins the exons and leaves a freed looped intron structure to leave, along with the U2, U5, and U6 snRNPs[11], as seen in (**Fig. 1.1 B**).

The end result of splicing is the deposition of a set of proteins along the joined boundaries of the 5' and 3' splice sites. These proteins are called the Exon Junction Complex (EJC) and are located usually 24 nt upstream of the 5'-3' boundary[12]. The minimal components of the EJC consist of the DEAD-box protein eIF4AIII, to which a series of other proteins (Y14, Magoh, and MLN51) bind to clamp it in place on the transcript[13]. The presence of the EJC, and by extension, the act of splicing is essential for mRNP export. In metazoans, the EJC is where export machinery is recruited, and so unspliced transcripts are very poorly exported[14,15]. The deposition of the EJC temporally separates the early stages of transcription initiation from the later stages of processing and export by signaling the completion of splicing.

The EJC also plays a role in quality control. It forms the center of nonsense-mediated decay, in which surveillance machinery in the form of the SURF complex can recognize premature stop codons based on their aberrant presence upstream of an EJC. This allows the cell to degrade faulty transcripts before they are translated[16].

1.2.4 – 3' End Processing

Before the transcript can be exported however, termination and polyadenylation must occur. In brief, this occurs through Cleavage and Polyadenylation Specificity Factor (CPSF), Poly A Polymerase (PAP), Cleavage Stimulation Factor (CstF), and cleavage factors I and II (CFI and CFII). As pol II escapes the promoter, its CTD becomes phosphorylated at serine 2. Subsequently, this allows cleavage factor 1A to bind the CTD, where it is in contact with transcripts exiting. As shown in (**Fig. 1.1 C**), when it recognizes a polyadenylation signal twenty nucleotides upstream of the cut site it enables cleavage[17,18]. The endonuclease responsible for cutting the transcript free is the CPSF73 subunit of CPSF[19]. After cleavage, other factors associated with phospho-serine 2 CTD eventually lead to transcription termination[20]. From the new 3' end poly-A polymerase then adds on average 250 adenine residues[11] (**Fig. 1.1 D**). Polyadenylation of transcripts has been shown to be a prerequisite for export, and improper 3' end processing causes transcripts to be retained and degraded. Concomitant with polyadenylation is the binding of PABPN1, which has sequence specificity to poly-A sequences and is a marker for complete 3' processing and subsequent export from the nucleus[21] (**Fig. 1.1 E**).

1.3 – The TREX Complex

1.3.1 – Overview

While the pathways above describe mRNA maturation, the export of the mRNP into the cytoplasm occurs through the Transcription Export (TREX) complex. TREX is assembled sequentially. The complete TREX complex consists of ALY, UAP56, and the THO complex, of which the THO complex is the first to be recruited[22–25]. TREX is also known to interact with some of the previously described mRNA maturation factors. At the 3' end, Pcf11 is a critical transcription termination factor in yeast[26], and has been shown to directly interact with Yra1, the ALY orthologue[27,28]. While at the 5' end, depletion of PABPN1 significantly reduced ALY localization to the 3' ends of mRNPs, further indicative of association, although not direct interaction, between export factors and 3' maturation[29]. THO has also been shown to directly interact with Z3CH14, a poly-A binding protein and orthologue of yeast Nab2. In coordination, depletion of Z3CH14 or individual THO subunits led to an increase in transcript polyadenylation[30]. Overall, TREX association with both 5' and 3' factors shows a physical association between steps of mRNP processing and implies a possible coordination that ensures that all maturation steps have occurred before export[2]. What is important to note though, is that the exact coordination between mRNA processing at the 3' and 5' ends and mRNA export by TREX is still unknown. Aside from splicing, it is unclear if the timing of different steps of TREX assembly and disassembly, shown in **Fig. 1.2,** is dictated by direct interaction with 3' and 5' machinery, and a complete and holistic view of coordinated processing and export is still lacking.

1.3.2 – THO Binding

THO is the first TREX component which is recruited[2]. THO is a large multi-subunit complex consisting of hTho2, hHpr1, fSAP79, fSAP35, and fSAP24[2,31,32]. Although no structure exists of the entire complex from metazoans, the structure has been determined for yeast THO in complex with Sub2, the yeast homologue of UAP56, which shows an overall elongated boomerang-shaped structure with Sub2 sitting in the interior of the bend[32].

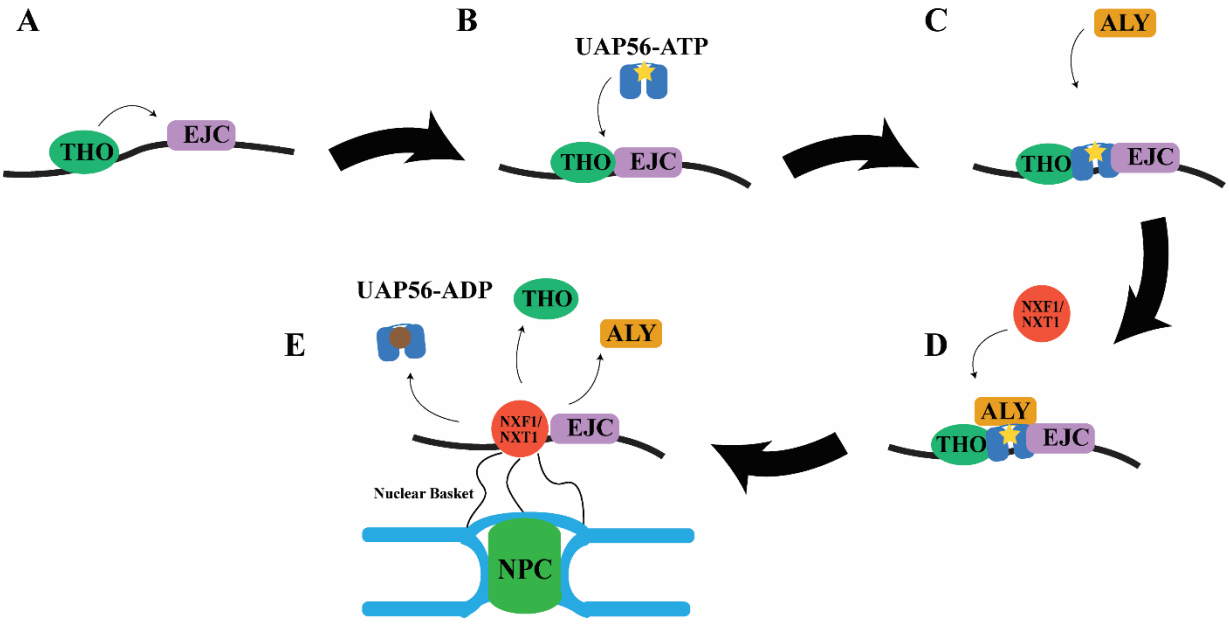

Figure 1.2 – Assembly of the TREX Complex

A) THO associated with the Pol II CTD recognizes and binds to a deposited EJC after splicing is complete. **B)** UAP56 with ATP bound associates with the EJC through interaction with THO. ALY is recruited by UAP56 through interaction with its C-terminus. The full TREX complex is assembled. **C)** The export receptor NXF1/NXT1 binds to ALY. This activates UAP56 ATPase activity and NXF1/NXT1 is deposited on ALY, while RNA is thought to be handed off from UAP56 to ALY. **D)** NXF1/NXT1 displaces the ADP-bound UAP56 which is the first component to leave TREX. **E)** At an uncertain point before binding to the NPC, THO and ALY also leave TREX. The export receptor is left bound to the RNA and the EJC. The mRNP is now mature and docks with the basket of the nuclear pore complex through NXF1/NXT1 interaction with Nup proteins as it undergoes biased diffusion out into the cytoplasm.

In yeast, THO binds directly to the poll II phosphorylated CTD and is one of the first complexes present during transcription[33,34]. In metazoans however, THO recruitment, as well as assembly of the rest of the TREX complex is recruited to the EJC[15,35,36] (**Fig. 1.2 A**). This difference reflects the fact that the vast majority of metazoan genes undergo splicing, while the vast majority of yeast genes do not, and so acts as a safeguard against premature export in metazoans. The structural difference that accounts for the different recruitment of THO between yeast and metazoans is not well understood.

While THO is essential for mRNA processing, a lot of the dynamics of its association with the other TREX components are still not well understood either. It is not exported to the cytoplasm, yet its association with ALY is required for binding to the export receptor NXF1[35,37]. So, it is unclear when THO leaves the mRNP[2].

1.3.3 – UAP56 Recruitment

THO, once recruited to the mRNP, then recruits UAP56 through its hHpr1 subunit. THO stimulates the ATPase activity of UAP56 by stabilizing a half open conformation which promotes ATP and RNA binding between the two RecA domains[32] (see **Fig. 1.2 B**).

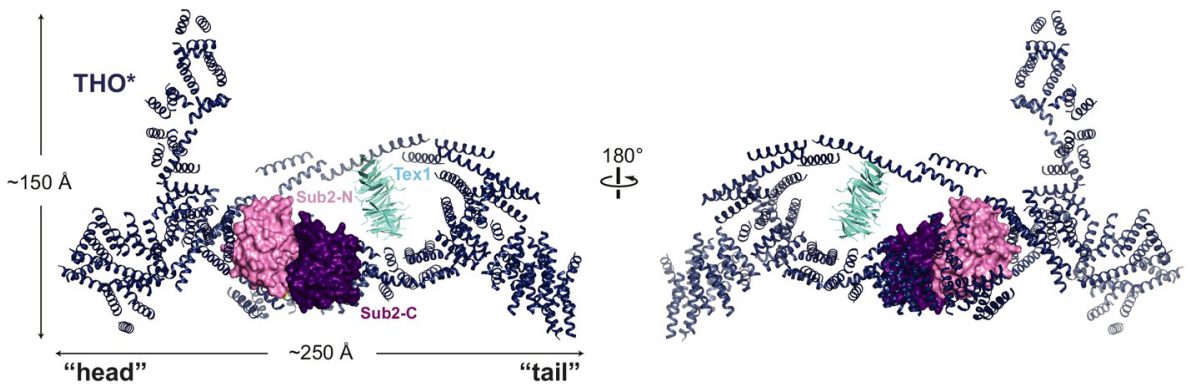

Figure 1.3 – Structure of Yeast THO bound to UAP56 Homolog Sub2

Crystal structure of the THO complex from *S. cerevisiae* with Tex1 protein (a part of TREX in yeast) from *S. bayanus* and the yeast UAP56 homologue Sub2 (PDB: 5SUQ). Resolution could not allow for the identification of individual THO subunits but did show an elongated and bent structure with two lobes. Notably, THO binding induces a primed half-open conformation of Sub2 to increase its ATPase activity. Figure taken from Ren *et al.* 2017[32].

As a central player in this project, the intimate details of UAP56 will be discussed further in a later section. Its principal role in mRNA export is the recruitment of ALY. It is likely that UAP56 binds to THO first, then is primed to bind to the mRNA by this THO-stabilized conformation, then RNA binding allows the hydrolysis of ATP thought to be necessary for later steps[2] (**Fig. 1.3**).

1.3.4 – ALY/REF Recruitment

The last component of TREX to be recruited is ALY, sometimes called REF. ALY recruits NXF1/NXT1, which in turn, acts as the adaptor with the nuclear pore. Unlike the other members of the TREX complex, ALY is reported to be dispensable for transcript export, at least in Drosophila and C. elegans. However even in these studies, ALY knockdown produced some

level of mRNP export defect. This suggests that another factor may be functionally redundant with it[38,39]. This would be consistent with reports that Thoc5, part of the THO complex, is capable of binding to NXF1 directly[40,41].

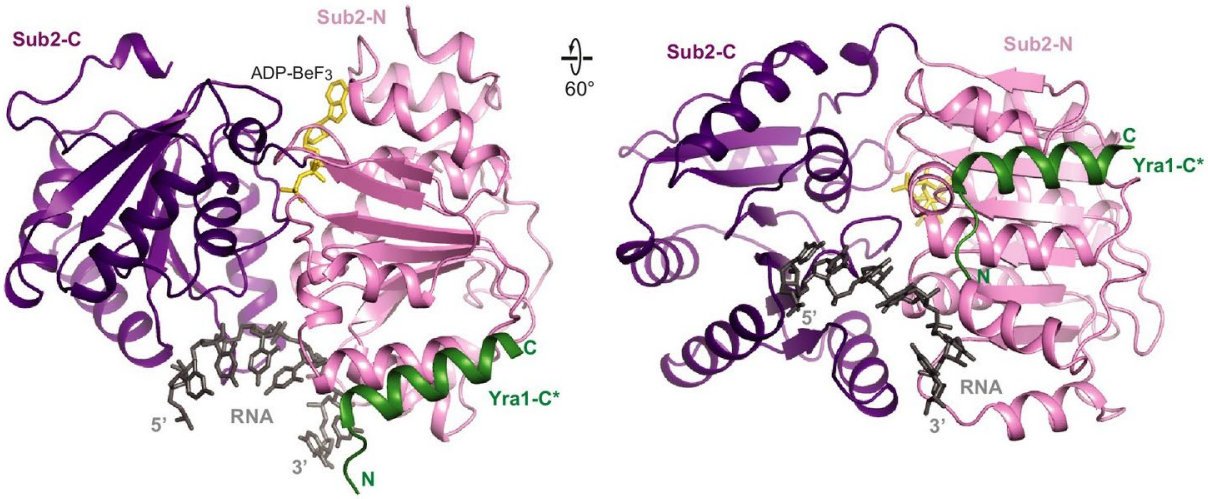

Figure 1.4 – Structure of Sub2 with Yra1

The crystal structure of the UAP56 yeast homologue Sub2 with the C-terminus of the ALY homologue Yra1 (PDB: 5SUP), demonstrating Yra1 manner of recruitment. Sub2 seen here is in the closed, pre-catalytic conformation with the non-hydrolysable ATP analog ADP-BeF_3, demonstrating the common DEAD-box mechanism of inducing a kink in the bound RNA between the Rec-A domains which does not allow duplex formation. Yra1 binding stimulates the ATPase activity of Sub2 to allow release for the bound strand. Figure taken from Ren *et al.* 2017[32].

ALY directly interacts with transcripts as part of its function, yet its recruitment is mediated through protein-protein interaction with UAP56 as well as THO (**Fig. 1.2 C**). The C-box domain, the C-terminal-most element of the yeast ALY homologue Yra1, was crystallized with Sub2, the yeast UAP56 homologue (**Fig. 1.4**). It was seen to be packed against the Sub2 N-terminal RecA domain, with Sub2 in a closed RNA-bound conformation. This suggests that loading of ALY

onto RNA requires the ATPase activity of UAP56. In support of this, ALY stimulates ATP hydrolysis by UAP56, a mechanism conserved in yeast[32,42,43].

Although it forms a complex with the receptor that delivers the mRNP to the nuclear pore, ALY itself does not leave the nucleus. In yeast, the ALY homolog Yra1 is removed as a result of ubiquitination by the protein Tom1[2,44]. It is unclear if this same mechanism holds in metazoans.

1.3.5 – NXF1/NXT1 Recruitment

Once integrated into the TREX complex, ALY will bind to the export receptor NXF1/NXT1 at its N-terminal RNA binding domains through an N-terminal arginine rich motif on ALY[37,45–49] (**Fig. 1.2 D**).This recruitment of NXF1/NXT1 is concomitant with the dissociation of UAP56 from the TREX complex, as demonstrated in yeast[24]. UAP56 and NXF1 have overlapping binding regions on the N terminus of ALY and are observed to be mutually exclusive[46]. NXF1/NXT1 makes protein-protein contacts with the TREX complex, but also makes direct contact with RNA through its leucine-rich repeat, RNA recognition motif, and NTF2L-like domains. It was found that NXF1 also competes with RNA for access to ALY, but that ALY significantly enhanced the NXF1/NXT1 complex affinity for RNA, leading to a model whereby ALY hands off the RNA to the NXF1/NXT1 complex[48] in a sequence independent manner[50,51]. In line with the aforementioned dispensability of ALY, two other proteins, 9G8 and SRp20, were shown to interact with NXF1 in a structurally similar manner[52].

1.3.6 – Export Through the Nuclear Pore Complex

To leave the nucleus, the NXF1/NXT1-bound mRNPs must travel through the NPC. For this, the export receptor interacts with Nup proteins[53]. NXF1 serves as the principle adaptor between the mRNP and the nuclear pore through a low-affinity, high avidity interaction with the disordered hydrophobic tails of Nup proteins in the nucleoplasmic basket of the NPC[53] (**Fig. 1.2 E**). Notably, unlike the typical karyopherin pathways used to import and export proteins from the nucleus, bulk mRNA export does not rely on a Ran GTP gradient[31,54,55]. The center of the NPC channel is filled with more unstructured hydrophobic Nup tails which are enriched in repeats of phenylalanine and glycine, or glycine-leucine-phenylalanine-glycine. This forms a hydrophobic barrier that effectively makes a gel-like phase through the center of the pore[56]. It is the favorable interaction of the export receptor with this milieu of hydrophobic residues that allows the mRNP to pass through the pore[57]. Studies of a very large mRNP called the Balbiani ring indicate that initially this contact occurs with the 5' end of the mRNP entering the pore first[58].

The directionality of movement through the pore is not any form of active pushing, but rather biased Brownian motion that presents an energetic barrier towards the mRNP sliding backwards[31,59]. This bias to diffusion is provided by the removal of NXF1/NXT1 by DBP5 and its activator Gle1. This occurs as soon as the mRNP emerges into the cytoplasm, as DBP5/Gle1 are bound to Nups on the cytoplasmic face of the NPC. The absence of the NXF1/NXT1 to interact with the Nups means the protruding portion of the mRNP will not move back into the pore easily[31,60–63]. After this point nuclear export is complete and the various processes involved in translation can begin.

1.4 – UAP56 in Detail

1.4.1 – Overview

UAP56 belongs to a family of proteins called DEAD box helicases, so named after the amino acids in their catalytic signature motif. The prototypical member of this family, and the most studied is eIF4A[64]. The core function of these helicases is the unwinding of double stranded RNA through ATP hydrolysis. They will bind double stranded RNA through the phosphate backbone without sequence specificity, and separate the two strands, releasing them independently. In UAP56, the N-terminal RecA domain (RecA-N) consists of residues 44-251 and the C-terminal domain (RecA-C) of residues 261-428. In between is a short flexible linker with poor sequence conservation between DEAD box helicases (**Fig. 1.5 A**).

1.4.2 – UAP56 Structure Comparison

Several full length structures of DEAD box helicases have been crystallized[32,65–67]. In all observed cases, the RecA domains have almost identical folds, however the relative orientations of RecA-N and RecA-C varies wildly without bound substrate, and they do not interact. In structures with RNA and nucleotide, both RecA domains contribute to binding the substrates through conserved motifs. The RecA domains consists of a central parallel beta sheet (six strands for RecA-C, seven strands for RecA-N) with alpha helices on both sides of the strand that run from the tail of one strand to the head of the next. RecA-N has eight helices and RecA-C has five.

The DEAD box helicase Vasa, crystallized in a closed conformation, first revealed that unwinding is accomplished by inducing a sharp kink in the RNA that forces the strands apart[65].

Double stranded RNA is bound in a sequence-independent manner through its phosphate backbone by motifs Ia and Ib (UAP56 a.a. 121-126 and 172-178 respectively) from one side by RecA-N while motifs IVa and V (UAP56 a.a. 324-327 and 341-347 respectively) engage the phosphate backbone from the other (**Fig. 1.5 B** and **C**). Contact occurs only on one strand of RNA, with the wedge helix in RecA-N (UAP56 a.a. 173-182) pressing against the phosphate backbone and bending the bound RNA strand in such a way that it cannot base-pair with the other strand without steric clash with the protein. The bound strand is retained, while the unbound strand floats off into solution, or otherwise disengages in the case of hairpins, and this prevents reannealing.

Both RecA domains also engage ATP together. In RecA-N, a universally conserved glutamine (Q72 in UAP56) engages the adenine base. Motif I (a.a. 89-96) positions the triphosphate moiety. Motif II, the namesake DEAD box, is actually DECD in UAP56, and coordinates the catalytic Magnesium ion between beta and gamma phosphates, as well as coordinating these phosphates along with Motif I. RecA-C contributes motif Va (a.a..348-352) and VI (a.a. 373-380) which coordinate the triphosphate from the other side and allow catalysis when the two domains come together (**Fig. 1.5 B** and **C**). RecA-C engages in a conserved aromatic stacking with the adenosine base. In the UAP56/ADP structure this stacking with F381 is lost as the RecA-C rotates outward, as are all the RecA-C contacts with the ADP phosphate groups[32,66]. While ATP binding is required for strand separation, its hydrolysis is not. Rather, hydrolysis is required for the release of the bound strand and for the enzyme to reset[64]. Unlike most helicases, DEAD box helicases have no real processivity as they do not move along a duplex strand, and typically unwind a few bases before dissociating[68].

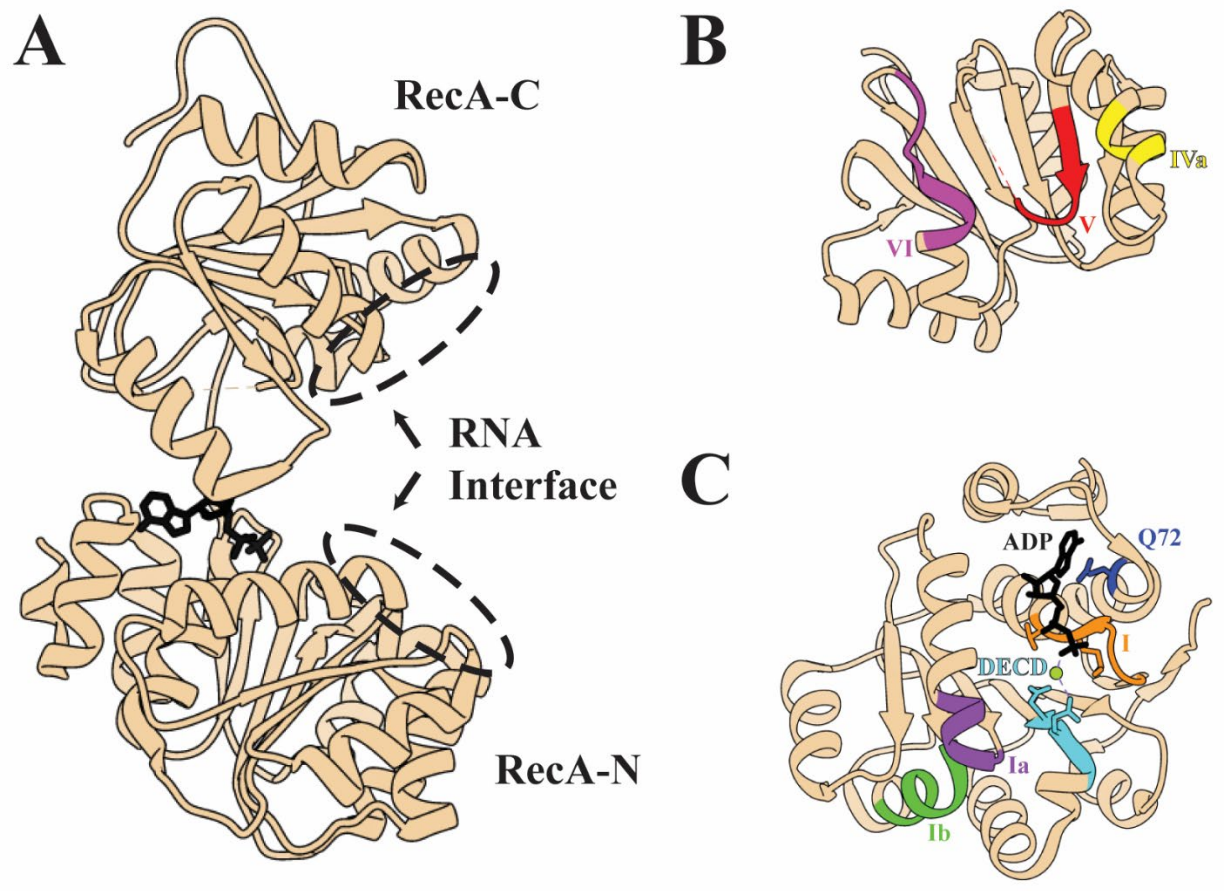

Figure 1.5 – The Structure of UAP56

A) A ribbon diagram of UAP56 bound to ADP from PDB: 1XTJ[66]. Note that the N-terminal element is not present in this structure. **B)** A view of the same model looking up at the C-terminal RecA domain showing conserved DEAD-box motifs. **C)** A 180° view looking down at the N-terminal RecA domain from above, also highlighting conserved motifs. Figure generated with UCSF Chimera[69].

Some DEAD-box helicases may also have accessory domains which complement the function of the core RecA domains, often targeting them to specific sequences or structural features.

The bacterial proteins Hera and DbpA each additional C-terminal RNA binding domains involved in recognition of specific ribosomal RNA[64,70,71]. The yeast protein Mss116 as well as the protozoan CYT-19 also have C-terminal factors involved in enhancing the RecA core

helicase activity, as well as being implicated in binding double stranded RNA separately[72,73]. While these proteins have functions for specific RNAs, UAP56 has no additional bona fide domains and does not exhibit any sequence specificity on its own.

However, for UAP56 as well as others such as Dbp5, DDX19, and DDX25 there is an additional N-terminal sequence before RecA-N. This has some predicted helical propensity for UAP56 but is largely unstructured and unresolved in the crystal structure before residue 45[66]. Interestingly though, an N-terminal extension is critical for auto-inhibitory functions in DDX19, and is shown wedged between the RecA domains in a crystal structure, but is displaced by RNA and nucleotide[74], and it is believed to operate this way for the yeast DBP5 ortholog as well[75]. A structurally homologous N-terminus to DDX19 was shown to autoregulate the DEAD box protein DDX25, suggesting N-terminal autoregulation may be a common feature[76]. While UAP56 N-terminal extension (NTE) is shorter than that of DDX19 and has low homology, it too has demonstrated autoregulatory function. However, deletion of the N-terminal 43 a.a. of UAP56, in contrast, approximately halved the ATPase activity of UAP56p,o, suggesting an auto enhancing rather than auto inhibitory role.

Dbp5 was crystalized with Gle1 and inositol hexaphosphate which showed Gle1 interaction with the N-terminus Dbp5 that prevents helical insertion between the RecA domains and structurally explains how Gle1 and inositol hexaphosphate can act as cofactors that increase Dbp5 ATPase activity by over ten-fold in vitro. Given that protein regulators can alter the activity of related DEAD box helicases, a similar regulation could possibly exist in UAP56 or Sub2 though the NTE. Indeed, even though only residues 62 onward are resolved in the THO Sub2 complex, they are firmly packed against THO, likely indicating interaction with the entire Sub2 NTE. Correspondingly, THO increases Sub2's ATPase activity over 3-fold[32].

1.4.3 – Other UAP56 Functions

While not discussed in this work, it is important to note that UAP56 has other cellular functions that just its role in the TREX complex. One of the primary and critical function of UAP56 outside of the TREX complex is as an essential splicing factor. UAP56 binds to the early splicing factor U2AF65 which loads the branch point binding snRNP U2, forming the A complex. It was shows that the ATPase activity of UAP56 is what allows U2 to be deposited on the transcript RNA. Indeed, UAP56 gets its name from this function, as it was originally described as 56 kDa U2AF65 Associated Protein. This was originally discovered by a yeast two-hybrid screen for U2AF binding factors, and it was shown that immunodepletion of UAP56 from cell extract prevented splicing in in vitro assays[77,78]. Later studies showed its ATPase function was required at later steps in splicing, and that its helicase activity unwinds duplexed RNA from the U4 and U6 snRNPs as well[79].

Another moonlighting role of UAP56 is in the generation of piRNA. These small transcripts have several functions, most notable among them is the suppression of transposable elements by a mechanism similar to RNA interference seen for siRNA and miRNA. Despite similar functions, the biogenesis of piRNA is incompletely understood[80]. UAP56 together with THO has been shown to play a critical role in their transcription and export into the cytoplasm, however[81–83]. This multifunctionality of UAP56 should not be surprising since it has no apparent sequence specificity and will bind any free RNA. Such a broad substrate range could mean that there are a myriad of RNA-directed processes across different pathways to which UAP56 is essential or acts on redundantly with other helicases, which have not been ascribed to it yet.

1.5 – The Influenza Virus

1.5.1 – Influenza Overview

Having described the host, we now turn our attention to the second aspect of this project, that of the influenza virus. Below is reviewed the basic biology of influenza to frame the context and scope of the interaction we are studying and its overlap with the host processes mRNA export processes reviewed above.

As a pathogen, influenza represents a major threat to worldwide public health. Seasonal epidemics cause millions of cases and about 500,000 deaths annually[84]. In addition, pandemics of more virulent strains also occur on a sporadic basis. The most infamous of these was the 1918 "Spanish" influenza, which killed an estimated 30 million people worldwide[85]. There is also the ever-present threat of the very highly pathogenic zoonotic strains of the virus, or "bird flu" as it is commonly known. Certain stains are endemic to populations of animals, notably waterfowl and swine, and can cause a very high mortality rate when human transmission occurs. Should certain adaptive mutations occur that allow for easier human-to-human transmissibility in these strains, the death toll could be massive. To understand where this project and the influenza nucleoprotein fits in the larger picture of the biology of the virus, an overview of the viral lifecycle is presented here.

There are four principal types of influenza virus, A-D. While influenza C and D are rarely seen in humans, A and B are more common. The A virus mutates considerably quicker and represents the majority of all cases in humans. Additionally, it is the only antigenic type to have caused pandemics[86]. Influenza A strains are classified according to two of their proteins, hemagglutinin and neuraminidase, with a number denoting the order of their discovery. The combination of the

specific variants of these two, with H indicating hemagglutinin and N denoting neuraminidase (e.g. H5N1) defines the classification group[87]. Influenza is a single stranded negative sense RNA virus whose genome is divided into eight segments. Each segment encodes an open reading frame for a viral protein, although some segments encode for multiple, for a total of ten essential viral proteins, although other minor proteins also exist in many strains[88]. These are hemagglutinin (HA), neuraminidase (NA), matrix protein (M1), the matrix ion channel (M2), the non-structural proteins NS1 and NS2, the two basic subunits of the viral polymerase (PB1 and PB2), the acidic subunit of the viral polymerase (PA), and the nucleoprotein (NP), which are summarized in **Fig. 1.6**.

1.5.2 – Influenza Virion Structure

The overall physical architecture of the individual virion was first characterized by negative stain EM studies done in the late 1960s[89,90], where it was seen that individual virions were round membrane-enveloped particles typically about 100 nm in diameter with proteinaceous spikes consisting of HA and NA along the exterior embedded in the membrane with a protein coat of M1 on the interior forming a sphere. The matrix protein M1 provides the virus with structural support. It forms a coat underneath the envelope membrane which encloses the viral contents in a sphere. Tomography revealed that the interior of the capsid contains helical-looking protein

Protein	Abv.	Function
Matrix 1	M1	Forms interior capsid of virion. Binds to vRNPs to export them. Anchors vRNPs to capsid.
Matrix 2	M2	Transmembrane protein in viral envelope. Acidifies virion durring entry. Promotes membrane bending durring budding.
Polymerase Basic Subunit 1	PB1	Catalytic subunit of the viral RNA-dependent RNA polymerase. Carries out both transcription and replication.
Polymerase Basic Subunit 2	PB2	Integrally bound with PB1. Binds 5' UTRs of host mRNA and host Pol II CTD for cap-snatching. Regulates PB1.
Polymerase Acidic Subunit	PA	Performs endonuclease function which cleaves host mRNA 5' UTR during cap-snatching. Binds vRNA 5' panhandle.
Non-Structural Protein 1	NS1	Blocks host mRNA export and polyadenylation. Antagonizes innate imune factors in the cytoplasm.
Nuclear Export Protein	NEP	Allows nuclear export of progeny vRNP through CRM1 pathway. May regulate polymerase.
Neuraminidase	NA	Transmembrane protein at viral envelope. Cleaves sialic acid at host extra-cellular matrix to allow virion release.
Hemagglutinin	HA	Transmembrane protein at viral envelope. Binds sialic acid at host extra-cellular matrix to allow virion endocytosis.
Nucleoprotein	NP	Binds viral RNA without sequence specificity. gives structure to vRNPs. Interacts with M1, polymerase, and host factors.

Figure 1.6 – Summary of the Ten Essential Virally Encoded Proteins

filaments which contain the genetic material of the virus called viral ribonucleoproteins (vRNPs)[91]. M1 also interacts with the vRNPs, anchoring them to the capsid[90–92]. The vRNP and its functions are the area of influenza biology which are the focus of this project. In brief, they are the packaged genome segments, of which there are eight per virion. Each consists of a viral polymerase which binds the 5' and 3' ends of each of the single stranded RNA genome segments called viral RNA (vRNA). The remainder of the vRNA is coated by NP, which binds independently of sequence like beads on a string to form a double helical hairpin structure capped on one end by the polymerase. The vRNP is described in detail in a later section. The structure of the virion and the viral lifecycle are described below and illustrated in (**Fig. 1.7**).

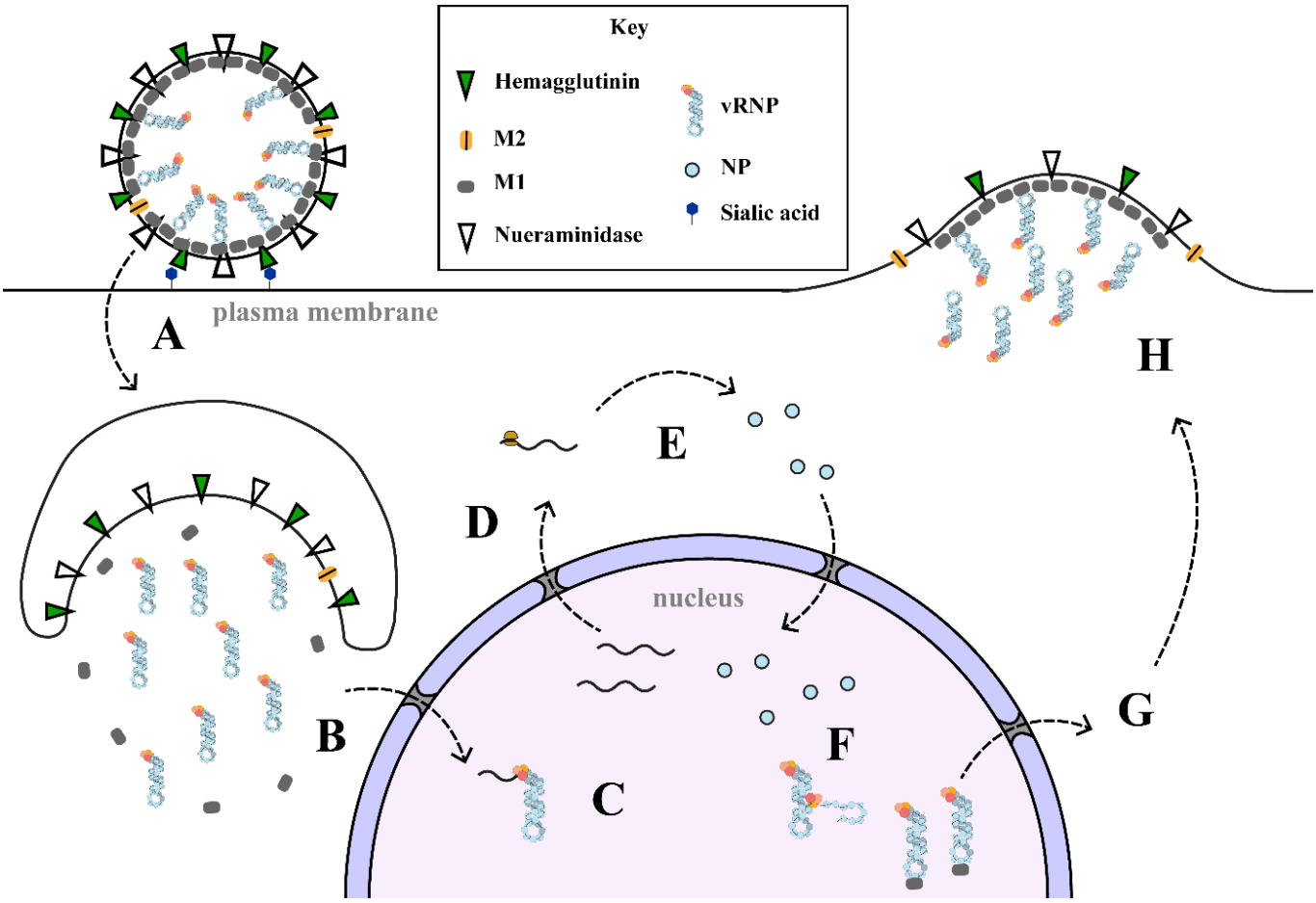

Figure 1.7 – The Simplified Influenza Lifecycle

A) A new virion adheres to the cell surface by binding of hemagglutinin to sialic acid, which triggers endocytosis. **B)** The early endosome acidifies and an influx of protons through the M2 channel into the virus causes disassembly of the M1 coat. Hemagglutinin causes membrane fusion and release of contents. vRNPs are imported into the nucleus by Importin alpha. **C)** Transcription of viral genes begins as the viral polymerase binds and cleaves 7-methyl guanosine caps from host genes to produce its own transcripts. **D)** Host transcripts are retained in the nucleus by NS1 sequestration of NXF1/NXT1, whereas viral mRNAs are selectively exported **E)** Translation of viral genes occurs. PA, PB1, PB2, NP, NS1, and NEP are imported back into the nucleus. **F)** Viral replication occurs as a switch between primer-dependent and primer-independent viral polymerase activities. It is thought that the switch occurs based on the availability of free NP within the nucleus. Replication occurs first to a positive sense cRNA then replication of cRNA produces another negative sense vRNA, which are bound by NP as they emerge from the polymerase. **G)** Influenza M1 protein bound to vRNPs acts as an adaptor for NEP and the Crm1 export pathway. **H)** vRNPs migrate to the plasma membrane where hemagglutinin and neuraminidase are incorporated. M1 forms a coat along the interior and together with M2 induce membrane curvature to bud off once all eight vRNPs are incorporated.

1.5.3 – The Viral Lifecycle

1.5.3.1 – Viral Entry

Infection begins when a virion makes contact with the extra-cellular surface of the host cell. HA on the virus surface binds sialic acid on the extracellular matrix of host lung epithelial cell, anchoring the virion for cellular entry. Cell surface binding triggers endocytosis by the host cell via either clathrin-mediated or macropinocytosis mechanisms (**Fig. 1.7 A**). The endocytosed vesicle matures into a late endosome as it migrates inward from the plasma membrane, which results in its acidification[93]. The interior of the capsid enclosed within the M1 protein coat is also acidified by way of M2 which acts as a transmembrane proton channel. Acidification of the viral interior results in the dissociation of the vRNPs from the M1 protein and the breakup of the coat, owing to a pH dependence of its self-association[94]. The dissociation of the vRNP from M1 is required for its later nuclear import[95]. This change in pH also causes a conformational change in HA that results in the fusion of the viral membrane to the endosome, which releases the viral contents into the cytoplasm[96] (**Fig. 1.7 B**). Of note, blocking the acidification of the viral capsid prevents any further progression of infection. M2 channel blockers like amantadine were once used as effective clinical treatments for influenza, before selective pressure led to M2 mutations which prevented drug binding[97]. The released vRNPs are imported into the nucleus through the nuclear pore complex in a Ran-GTP dependent manner[98,99]. It was shown that a non-conventional NLS of NP at the very N-terminus recruits importin alpha /beta to utilize the import pathway typical of most nucleo-cytoplasmic protein shuttling[100].

1.5.3.2 – Viral Transcription

Once imported into the nucleus, vRNPs begin producing mRNA transcripts of the protein encoded on that genome segment through the polymerase on the vRNP (**Fig. 1.7 C**). PA, PB1 and PB2 collectively form the viral polymerase, which is an RNA-dependent RNA polymerase. A large gamut of proteins, both viral and cellular, interact with the polymerase. PB1 forms the central subunit of the complex onto which PB2 and PA anchor as well as the active site of the polymerase.[101,102].

The transcriptase activity of viral polymerase depends on co-opting the normal host mRNA transcription activity of RNA polymerase II in order to snatch host-derived 5' transcript ends with the proper 7-methyl guanosine triphosphate cap[103–106] (**Fig. 1.8 A**). The binding of the host mRNA cap is performed through the PB2 subunit by a discrete domain in the middle of the protein[107,108]. To steal the cap, viral polymerase must bind to the host RNA polymerase II C-terminal domain as soon as it is activated by phosphorylation at serine 5, which it does at two sites on the surface of its PA subunit C-terminal domain[109]. The host mRNA is then cleaved between position 10-14[106]. This endonuclease function was previously thought to be performed by PB1[110], but it now known that PA performs this via its N-terminal domain[111]. The viral polymerase then completes a positive sense mRNA by moving along the vRNA, adding viral sequence to the 3' end of the stolen cap.

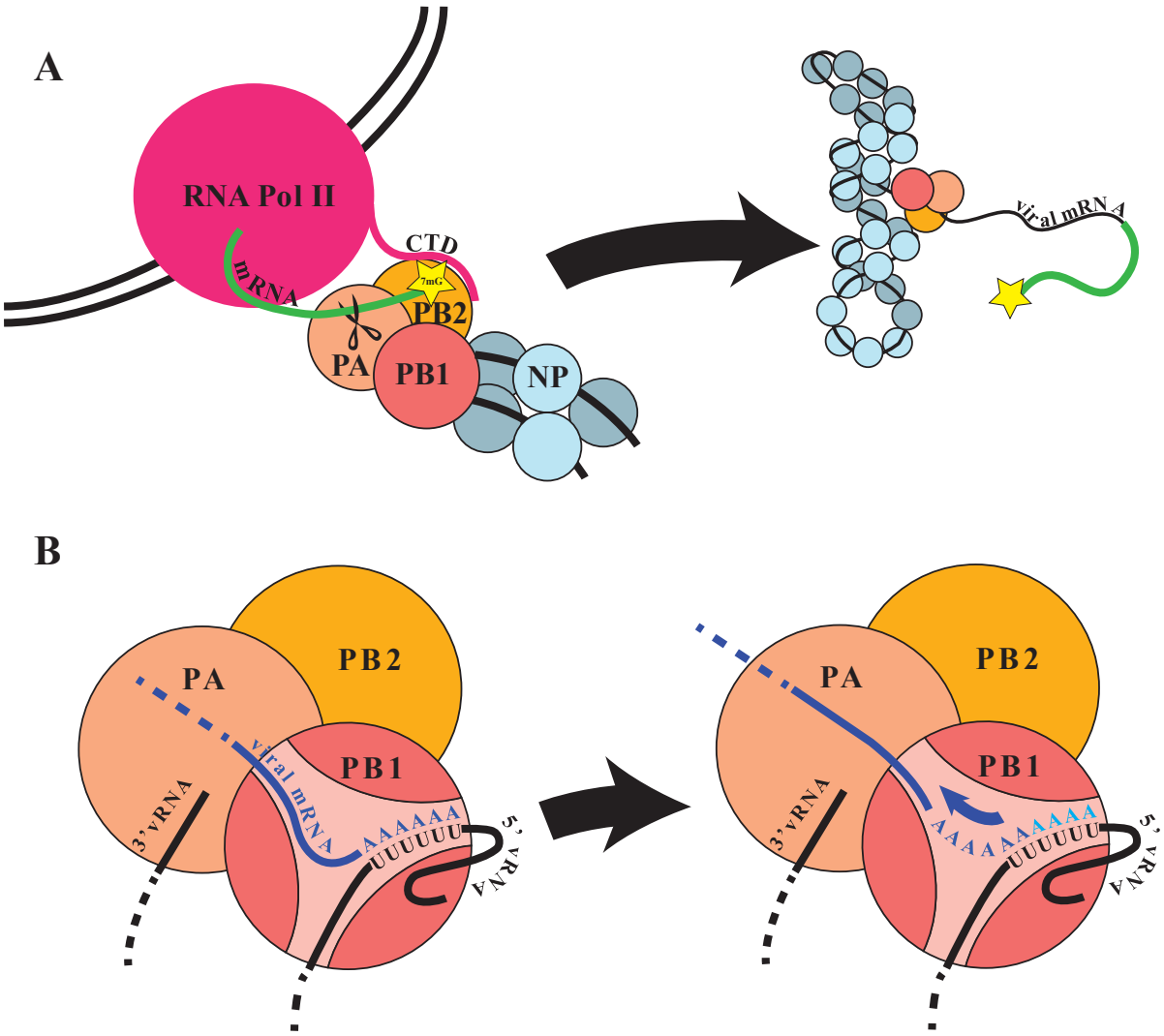

Figure 1.8 – Viral Transcription

A) schematic illustrating the viral mRNA transcription process through binding Pol II CTD and binding the 7-methyl-Guanine cap via PB2, and cleavage by PA, for subsequent elongation by PB1. **B)** when the polymerase reaches the 5' of the vRNA template, a polyuridine stretch causes the polymerase to stutter and re-transcribe additional adenosine stretches to polyadenylate viral mRNA independently of the host 3' maturation factors.

Like the host's mRNA, viral mRNA is also polyadenylated to avoid degradation and be properly exported. Unlike the comparatively sophisticated mechanism employed for polyadenylation of host transcripts, however viral mRNAs are polyadenylated when the viral polymerase encounters

a poly-uridine tract at the 5' end of the template. This causes it to stutter and slip, repeatedly copying this region to produce a poly-A tail on the transcript (**Fig. 1.8 B**)[112]. Notably, two of the influenza genome segments can undergo post-transcriptional splicing at nuclear speckles. Segment 7 is exported predominantly unspliced to encode the M1 protein, but a portion is spliced to produce mRNA for M2. Similarly, segment 8 produces NS1, but if spliced instead produces NEP[113]. Although it is incompletely understood how the interaction occurs, NXF1/NXT1 which is the exporter for cellular mRNA, is also used for viral mRNA export[112]. At the same time, export of cellular mRNA is blocked by NS1[114], abrogating the host's ability to respond with production of interferon-stimulated genes (**Fig. 1.7 D**).

1.5.3.3 – Viral Replication

When the influenza proteins are translated in the host cytoplasm, hemagglutinin, neuraminidase, and M2 are directed to the Golgi where they undergo the typical processing of host membrane proteins and make their way to the plasma membrane where they localize to lipid rafts[115]. M1, NS1, PA, PB1, PB2, and NP all have nuclear localization sequences and are imported into the nucleus once produced, while NEP is small enough to enter passively (**Fig. 1.7 E**). When enough structural proteins have been produced for virion assembly, vRNPs switch from production of mRNA by cap snatching to self-primed replication of the vRNA to a positive sense intermediate called cRNA. What factors trigger this is not completely understood, but the availability and nuclear accumulation of free NP is thought to play a large role, as it is required for binding progeny vRNPs, but not has not been found associated with viral mRNA[116]. It is also known to interact with the polymerase PB2 subunit, and this may shift it into an alternate polymerase

activity[117]. NEP has also been implicated in this role. In a cell-based study, plasmids for the minimal viral proteins for replication (PA, PB1, PB2 and NP) were transfected into cells and each remaining viral protein transfected in to test them individually for their ability to regulate levels of viral mRNA, cRNA, and vRNA. NEP co-expression altered polymerase activity by suppressing mRNA production and promoting replication to cRNA[118].

Unlike viral transcription, which is highly efficient because it is primed by cap-snatching, viral replication is considerably less efficient as it involves de novo synthesis without a true primer beginning at the 3' end of the vRNA to produce an exact positive sense copy of the genome (cRNA) without any caps or polyadenylation seen for viral mRNAs. This then repeats to make another negative sense vRNA from the cRNA complement[93]. A crystal structure of the polymerase complex in the pre-replication state (PDB:4WSA) led to a putative model for how de novo replication of either cRNA from vRNA or vRNA from cRNA is carried out (**Fig. 1.9**)[119]. It is thought to depend on a beta hairpin priming loop in PB1 (residues 641-657) based on functional knowledge and structural comparison to other viral RNA polymerases from bacteriophage[120] and hepatitis C [119,121]. For cRNA synthesis, this priming loop acts to stabilize the first incoming nucleotide, GTP, in a base pair with the second to last C at the 3' terminus. An ATP then comes in on the 5' side of this GTP in a rate limiting step to form the dinucleotide pppAgG. From this point elongation occurs, which presumably involves the repositioning of the priming helix as duplex RNA is extruded. The polymerase then replicates using the 3' terminus being fed in as a template. It has been proposed that vRNA synthesis occurs by the same manner but involving a repositioning of the pppApG dinucleotide to the last 3' template base after its initial formation at position 4 and 5[122–124]. It has also been proposed that in contrast to viral

transcription, replication must proceed by the transfer of the 3' terminus from one vRNP to a non-resident polymerase. This model is not universally accepted, however[125].

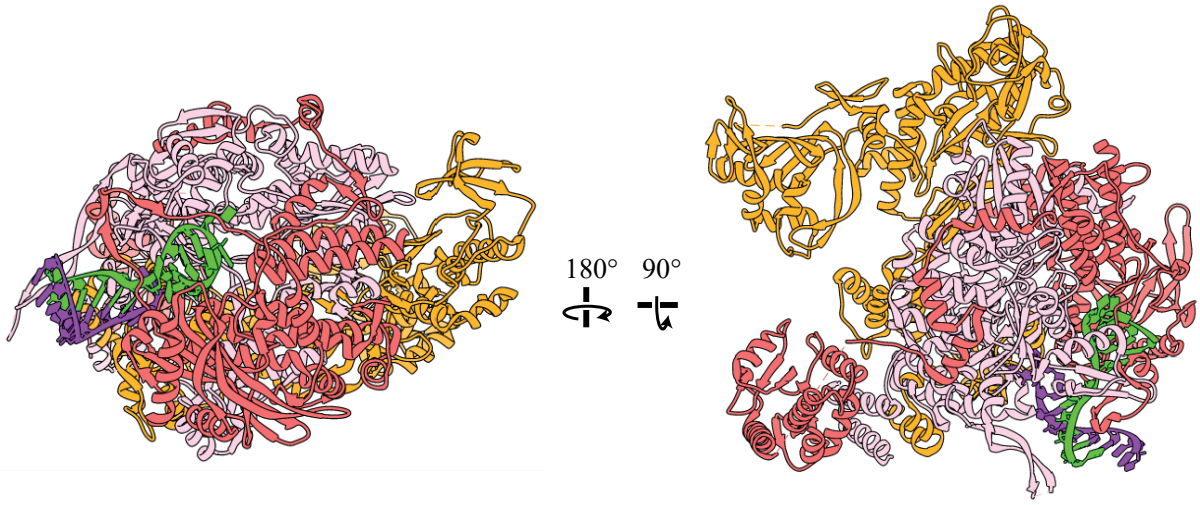

Figure 1.9 – Crystal Structure of Viral Polymerase in Pre-initiation State

PA is shown in salmon, PB1 in pink, PB2 in orange (PDB:.4WSA)[119]. The 5' end of the viral RNA in green forms a hairpin bound by PA and PB1 and forms a duplex with the 3' viral RNA terminus in purple. The endonuclease domain of PA and the cap binding domain of PB2 have rotated away from the template entrance and disengaged, promoting de novo replicative activity.

Newly synthesized cRNA and vRNA does not remain naked though. Upon emerging from the polymerase, NP covers the nascent RNA (**Fig. 1.7 F**). It is important to note that this process leads to abortive replication and premature termination in the absence of host factors, one of which is UAP56, to properly load NP onto new RNA. This aspect of viral replication forms the background context of this project within the scope of influenza biology and will be discussed extensively later.

1.5.3.4 – Export of vRNPs

Once produced, new vRNPs need to be exported from the nucleus to make their way to the plasma membrane. Whereas viral mRNA is exported by the NXF1/NXT1 pathway used by most cellular mRNA, vRNPs are exported by the CRM1 pathway. The M1 protein associates with assembled vRNPs through an interaction with NP mediated by RNA. As previously mentioned, M1 in complex with the vRNP prevents its nuclear import, so it is likely that M1 binding during its export serves as means to prevent re-importing vRNPs[95,126] (**Fig. 1.7 G**). M1 also serves as the intermediary to NEP. Through its ordered C-terminal domain NEP binds to M1 associated with a new vRNP while the disordered N-terminal domain contains two non-canonical nuclear export signals that are recognized by CRM1[127,128]. Crm1 then allows vRNP passage through the Nuclear Pore Complex (NPC) by binding to Ran-GTP and subsequent stimulation of hydrolysis to Ran-GDP by GTPase-activating proteins that causes CRM1 to release its cargo. Because this CRM1 pathway is essential for the virus but not for host bulk mRNA export, it has been a target for anti-viral drug development[129].

1.5.3.5 – Nascent Virion Assembly and Budding

Exported vRNPs then make their way to the apical plasma membrane by utilizing host cytoplasmic trafficking. The vRNP is able to bind to Rab11 and migrate with recycling endosomes as they move outward to the plasma membrane on microtubule tracks[116,130,131]. M1 also assembles at the apical membrane to begin forming what will become the capsid of a new virion. There is evidence to suggest that the termini of both HA and NA, which extend into the cytoplasm, are responsible for recruiting M1 to the interior of the membrane at the budding site

in high enough concentrations that it can multimerize to form the interior capsid[132]. Through their association with M1, vRNPs are thus recruited to the budding site as well. When all 8 vRNPs have assembled at the apical membrane outward budding takes place (**Fig. 1.7 H**). The budding process that produces virions from the apical membrane is thought to be a product of membrane curvature induced by HA and a curved net-like lattice formed under the membrane by M1. Notably, mutations in M1 can cause virions to become elongated and filamentous[133]. M2 has a role in viral budding in addition to viral entry. It contains a cytoplasmic helix with a CRAC motif which recognizes cholesterol and has been demonstrated to induce the membrane curvature necessary for viral budding by its binding, both in model membranes and in vivo[134,135]. Once budding is complete, the last step necessary to produce an infectious virion is cleavage by neuraminidase. Like HA, neuraminidase is a transmembrane protein which forms part of the viral coat. It is a glycosidic enzyme which cleaves the terminal sugar of the sialic acid chain to allow for virion release. Notably, the most common antiviral drugs used to treat influenza such as oseltamivir and related compounds are neuraminidase inhibitors, which act as sialic acid mimetics[136].

1.5.4 – NS1

As previously stated, the interest of the Ren lab in influenza is related to the way in which it hijacks the host mRNA processing and export pathways for its own use, and how this can be used to study those host pathways. Any discussion of virus-host interaction must address NS1, as its primary function is disruption of host transcription, and the study of NS1 has revealed great insights into mRNA export. The protein consists of an N-terminal RNA binding domain and a C-

terminal effector domain with a flexible linker in between. The NS1 protein performs many functions for the virus. Although, as the name implies, it is not a structural component of the virus, it is still considered essential. Without it, the virus is only capable of even attenuated growth in cell lines with knockouts of innate immune pathways.

One of the most critical functions of NS1 is the selective export of viral mRNA and simultaneous nuclear sequestration of host mRNA. To accomplish this, it blocks host access to the essential mRNA export factor NXF1/NXT1[137]. Recently, a crystal structure was published by the Ren lab showing that this occurs by direct binding to the leucine rich repeat and nucleoporin-binding domains of NXF1/NXT1 through its effector domain through two conserved residues, F103 and F138 (PDB: 6E5U)[114]. Binding in this manner, it acts as a mimetic and competitor to Nup98, a component of the nuclear pore complex to which NXF1/NXT1 interacts to export mRNA, and so blocks access of mature host mRNPs to the nuclear pore (see **Fig. 1.10**). Notably, retention of host transcripts was non-uniform and transcripts of interferon stimulated genes were enriched in the nuclear-retained RNA. This was demonstrated in vivo by showing that nuclear retention of cellular mRNA resulting from viral infection could be reversed by overexpression of either the full length NXF1 or $NXF1_{201-616}$, containing the leucine rich repeat and nucleoporin-binding domains. Curiously though, overexpression of the converse NXF1 construct, $NXF1_{1-200}$, had a dominant negative effect in influenza infected cells, exacerbating the nuclear retention further, while lacking this effect in mock-infected cells. In agreement with the crystal structure, infection using influenza with mutations in the residues critical for NXF1 binding, NS1 F103A/F138A did not show significant nuclear retention of host transcripts[114].

In line with its role in disruption of host mRNA export, NS1 also binds to CPSF30 through the same interface at F103, as revealed in an earlier crystal structure. CPSF30 is required for the

proper polyadenylation of host mRNAs so its disruption leads to their degradation as another mechanism to suppress host gene expression[138]. In a similar manner, host mRNA processing is disrupted by NS1 interaction with another polyadenylation factor PABII. Notably, this interaction has no effect on viral mRNA as polyadenylation is performed directly by the viral polymerase instead[139]. NS1 also plays a role in regulating the splicing of its own transcript into the NEP coding sequence as well as the transcript of M1[113].

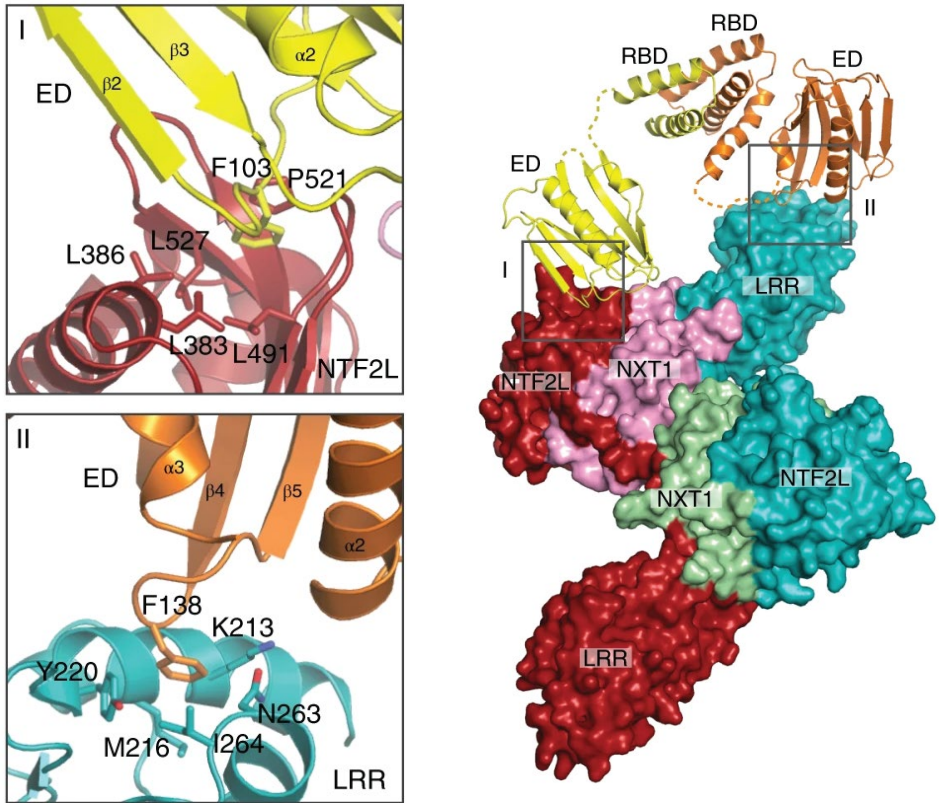

Figure 1.10 – Structure of NS1 with Host Export Receptor

Crystal structure of influenza NS1 in complex with the export receptor NXF1/NXT1 (PDB: 6E5U)[114]. Note the 2-to-2 stoichiometry with two NXF1/NXT1 heterodimers to two NS1 dimerized through their RNA binding domains. Opposite sides of the effector domains of NS1 through F103 and F138 interact with hydrophobic pockets in both the LRR and NTF2L domains of NXF1, shown in insets I and II. Figure taken from Zhang *et al*. 2019[114].

While not the focus of this lab, it also must be noted that NS1 has many functions in the cytoplasm to do with the evasion of the host innate immune system. Cells contain a surveillance system consisting of proteins called pattern recognition receptors. These monitor the cell for nucleic acids, proteins, or glycans not seen in a normal cellular environment which are indicative of a pathogen[140]. In the case of influenza, it is usually double stranded RNA. NS1 antagonizes several different factors involved in monitoring for viral RNA, including 2'-5' oligo-adenylate synthetase[141,142], RIG-I[143,144], TRIM-25[143], and protein kinase R[145].

1.6 – NP structure, Function, and Interactions

1.6.1 – Introduction

Having reviewed the general biology of influenza in its infection of host cells, we shift our attention now to the vRNP, and specifically NP. The replication of vRNP is the aspect of influenza biology at the center of our studies, and the context of the UAP56 NP interaction which we aim to characterize. To provide the background to understand this interaction, a general description of the vRNP, and an intimate description of NP are provided.

1.6.2 – General vRNP structure

The earliest structural information of vRNPs came from electron microscopy of sub-viral components separated by gradient ultracentrifugation, which revealed them to be flexible elongated structures of lengths between 50 and 130 nm consisting of a double helix with a loop

connecting the two helical strands at one end and globular mass later identified to be the polymerase at the other[89,90,146]. Treatment of both reconstituted and native vRNPs with RNase V1 revealed bound RNA is partially resistant to cleavage and indicated an arrangement which still contained dsRNA elements. This suggests that RNA bound to vRNPs can still retain secondary structural elements in at least some places[147].

At the same time, when vRNPs are treated with nuclease specific for single stranded RNA, vRNA was highly digested, indicating that the arrangement of NP within the vRNP is such that RNA is surface accessible, likely with the RNA wrapping around the outer edge of the helix[148,149]. The 3' and 5' ends of each viral genome segment are complementary to each other and highly conserved[150,151]. Mutational studies[152] as well as later structures of the polymerase complex revealed that the 5'-most end forms a hairpin and further down, the 3' end base pairs with it to form a double stranded stretch of about 14 nt to effectively link the two termini. It is this structure which the polymerase specifically recognizes and resides on at one end of the vRNP[119].

1.6.3 – NP Structure

Finally, there is the nucleoprotein (NP), the central focus of this dissertation. NP is a 56 KDa protein encoded on genome segment 5. The first crystal structure available for NP from the A/WSN/1933 strain (PDB: 2IQH)[153] revealed that the protein consists of a head and body domain, as well as a tail loop formed by residues 402 through 428. The head and tail domains form a contiguous crescent shaped body, with helices five through eleven as well as nineteen forming the smaller head domain and helices one to four as well as helices twelve through

seventeen forming the larger body (**Fig. 1.3 A**). Between the head and body is a large groove on one side opposite the tail loop. Although the head and body are connected through a series of loops, they are not mobile relative to each other[153]. At the N-terminus is a previously identified nonconventional nuclear localization sequence which the structure revealed to be sufficiently exposed to be functional[100,153]. NP sequence across strains of IAV shows fairly high conservation, with the most conserved parts being helices 6 and 7 as well as the head domain in general. The region surrounding the tail loop and helices eighteen and nineteen which follow it are the least conserved parts of the structure[146]. This is likely the result of their unimportance to structural stability, as most of the tail loop is unresolved in the first published crystal structure. A mutant made in the tail loop at R416 was crystallized with the tail loop in a different orientation relative to the body[154]. This indicates a high degree of flexibility within the loops leading to the tail loop. This flexibility very likely plays a function in the tail loop-dependent multimerization, elaborated on below, and allows for the flexibility in the angle with which each NP interfaces with its neighbor which accommodates any number of subunits to form either rings or a helical arrangement.

The principle function of NP is to provide protection and structure for the influenza genome. NP binds RNA in a sequence independent manner through contacts made between the positively charged side chains which line the cleft surface and the phosphate backbone of RNA (**Fig 1.11 B**). Because binding occurs through the phosphate backbone, NP is also capable of binding DNA[155]. In vivo however, it is almost exclusively bound to the negative sense vRNA or the positive sense cRNA replication intermediate[156]. As part of the formation of vRNPs, NP readily oligomerizes[155]. The primary interaction driving the oligomerization occurs between the tail loop and a crypt formed between several of the interior beta sheets, with critical salt bridges between

R416A and E339 and between R422 and E449[153]. Each NP inserts the tail loop into the crypt of its neighbor and polymerization occurs to form circular structures of varying sizes in vitro, even in the absence of RNA. In vivo however, the RNA binding and self-associative properties of NP lead to its accumulation onto cRNA and vRNA as they are being replicated by the influenza polymerase[116].

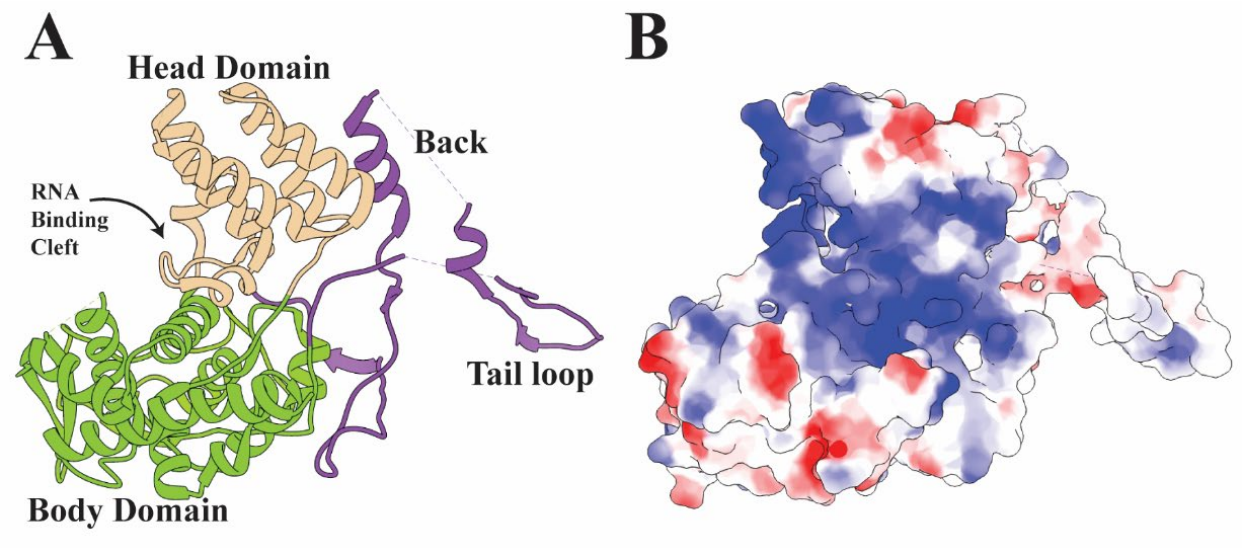

Figure 1.11 – The Structure of NP

A) A ribbon diagram of NP from A/WSN/1933 with domains highlighted (PDB: 2IQH)[153]. **B)** The same structure with surface electrostatics. Blue is negative and red is positive charge. Figure generated with UCSF Chimera[69].

1.6.4 – NP Phosphorylation

While not investigated in this study, it is important to note that NP can be phosphorylated[116,157]. Mass spectrometry-based phosphoproteomics showed NP to be the most heavily phosphorylated protein in influenza. Residues 3, 9, 78, 165, 188, 296, 377, 402, 457, and 472 have been

identified across various different strains as subject to phosphorylation to various extents[158–162]. Notably, the exact pattern of phosphorylated residues depends on strain, and changes over the course of infection[163].

Control of phosphorylation and dephosphorylation of serine 165 was shown to regulate the ability of NP to form oligomers. Phosphomimetic S165E critically inhibits oligomerization of NP by disrupting the interface between subunits, producing NP which exists as monomers in vitro, and cannot produce virions in a reconstitution assay. The phospho-deletion mutant S165A, however also exhibited reduced titer, suggesting that phosphorylation at this residue may be used to regulate NP oligomeric state, which changes from early to late infection[161].

Threonine 188 was recently identified as an important phosphorylation site. It is part of one of NP's nuclear export signals and mutation to a phosphomimetic glutamate led to the permanent nuclear retention of NP, as well as reduced replication and transcription by the polymerase which could not create infectious progeny using a plasmid-based viral reconstitution assay. The phospho-deletion mutant T188A had comparatively mild perturbation. While purely structural perturbation of NP itself can't be totally ruled out, this seems to indicate that phosphorylation is at least one mechanism by which the nucleocytoplasmic trafficking of NP is regulated as part of coordinating its export during the viral life cycle[160]. In a similar manner, the phosphorylation status of residue 78 as well as residue 9 within the nuclear localization sequence have been shown to affect nucleocytoplasmic shuttling of NP[159,162].

1.6.5 – NP-Polymerase Interaction

As part of its role in genome packaging, NP also regulates the activity of the viral polymerase. NP binding to nascently replicating RNA may prevent the polymerase stuttering seen at the end of viral mRNA transcription, in order to produce a perfect full-length complement. In doing so it then plays a critical role in governing the transcriptional versus replicative activity of the polymerase[148]. The first evidence for this interaction came from co-immunoprecipitation of NP with the viral polymerase from cell extract[164]. Later, minigenome replication assays looking at expression of firefly luciferase transcripts on cRNA demonstrated that R204, W207, and R208 were essential residues on NP for its association to PB2. Mutation severely abrogated the ability of the polymerase to produce the vRNA needed for subsequent reporter expression. Additionally, mutation at these sites prevented co-precipitation of polymerase with NP[165]. Another study identified that PB2 interacted with NP through both its N and C termini independently and in competition with those domains' interaction with PB1. In this manner, the presence of NP was postulated to disengage PB2 from PB1 and so shut off the polymerase's transcription function to engage in replication. This was supported by their finding that expression of PB2 N and C-terminal domains in isolation or overexpression of NP acted to reduce reporter gene expression as trans-dominant inhibitors[117]. Although this identified broad regions of interaction, it did not look at the mediating role of RNA. A later study however, showed that mutation of arginine 150 of NP in avian H5N1 strains led to a complete loss of all viral polymerase activity, while this effect was much less severe for human adapted H1N1 stains. In the avian strains, this could be rescued by the mutation of PB2 residue 627 from glutamate to lysine. They further demonstrated that NP was interacting with the 627 domain of PB2 specifically, and in competition to NP's RNA binding capacity. This position is known to be a key mutation in adaptation of avian stains

to mammalian hosts, and this implicates the polymerase-NP interaction in species adaptation of highly virulent strains. Interestingly though, these authors reported NP-polymerase interaction as being critical for both transcription and replication, rather than acting as a switch between the two[166]. Evidence is still ambiguous about the exact interface or interfaces by which NP can interact with polymerase as well as the functional effect this has on polymerase activity.

1.6.6 – vRNP Co-Packing

When new virions are being formed, the virus must ensure that each has all the genome segments contained. Electron microscopy studies have revealed that when leaving the cell and budding off into progeny virions, vRNPs consistently group together with each virion containing one of each genomic segment. Furthermore, these arrange in a specific '7 + 1' packing pattern inside the virion[167].

Recent mapping of the influenza genome by photo-crosslinking based high throughput sequencing methods[168,169] has shown that NP coverage along the length of vRNPs is non-uniform, with some spots being more bare than others. It was demonstrated that these more open regions contain RNA elements with more secondary structure, and it was hypothesized that these play a role in genome packing. The 3' and 5' ends of vRNAs were shown to be sparsely covered by NP, and these untranslated regions are known to be critical for proper packing, but without contributing any specificity to intersegment interactions[170]. Specific mutations in the head domain of NP can lead to disrupted coordination in genome packing, leading to the preferential incorporation of segment 3 over segment 7. Notably however, changes to 5' packing signal of segment 3 was able to reverse this, indicating a role for NP in packing between specific

segments[171,172]. A complete view of the mechanism by which vRNPs specifically pack to incorporate exactly one of each of the eight segments in nascent virions is still lacking.

1.6.7 – Other NP-Host Interactions

The Nagata group, which discovered the UAP56-NP interaction also identified a handful of other factors which promote viral replication using a single gene deletion library screen in yeast[173]. Prp18, a part of the eukaryotic spliceosome, was demonstrated to bind to the viral polymerase and NP directly, and in an RNA independent manner. In an in vitro system it was shown to facilitate the association of NP with RNA, as well as boost the anti-termination capabilities of NP in in vitro vRNP replication. In infected HeLa cells its knockdown was associated with decreased viral replication[174].

Another protein that was identified using the same screen and shown to have almost the same pattern of interaction is TAT-SF1. It too was shown to bind NP and promote its loading onto RNA[116,175]. TAT-SF1 is known to be a host factor in other viruses. Indeed, TAT-SF1 derives its name (tat specific factor 1) from its known role in regulating host transcription of integrated HIV viral genes[176]. The manner of its action, binding directly to NP and acting as a recruiting factor, as well as its common role with UAP56 acting in mRNA processing, suggests that it may be functionally redundant with UAP56.

Curiously, also identified as a host replication factor was the minichromosome maintenance complex (MCM). MCM's role in the host cell is as a DNA helicase involved in chromosome remodeling associated with transcriptional activation, and involved in cell division checkpoints[116,177]. MCM was shown to co-precipitate with PA specifically both in vitro and in

cells. Unlike other identified factors, MCM does not interact with NP, and seems to exclusively promote the elongation of cRNA as opposed to viral mRNA or vRNA. This was seen to be independent of its helicase activity, and it was postulated it may stabilize PA and cRNA in the initiation complex to prevent abortive replication[178,179]. If this tentative mechanism can be verified it would show that influenza has evolved multiple distinct routes of promoting its various polymerase activities through interactions with more than just NP and utilizing more than just the host splicing machinery.

The common theme among proteins that putatively regulate viral polymerase activity seems to be that they are connected to host mRNA maturation. TAT-SF1 and Prp18 are involved in splicing[180], while MCM has been reported to interact with Pol II CTD[181–183] much like influenza PA, while UAP56 interacts with the exon junction complex post-splicing[15]. It is of little surprise then that a mass spectrometry-based approach using tagged influenza polymerase expressed in mammalian cells was able to detect co-association of the influenza polymerase with a list of mRNA processing factors. This included polypyrimidine tract-binding protein-associated splicing factor (PSF), which was later demonstrated to enhance polyadenylation of viral[184] transcripts as well as UAP56 paralogs DDX3 and DDX5[185]. This raises the intriguing question of whether the larger host transcription and mRNA processing macro-assembly might be hijacked in concert, with the individual host subunits cooperatively promoting viral polymerase activity or if these interactions have evolved and function independently. There is currently no evidence for this, but hopefully this might be addressed as more studies are done, as there are certainly more interacting factors yet to be identified.

1.7 – The UAP56-NP Interaction

Why is UAP56 required by influenza for its replication, though? PA, PB1, PB2, NP, and nucleoside triphosphates should in theory be enough to allow viral replication as the viral polymerase carries out catalysis of all RNA used by the virus. It has long been known that vRNPs in isolation were not capable of replication. It was demonstrated that when vRNPs were pelleted from infected cell nuclear extract, the addition of the supernatant from that extract as well as ApG dinucleotide was necessary to replicate vRNA. This furthermore shut down viral mRNA transcription. The factor within the supernatant fraction responsible for this was identified as NP[186]. This experiment however, used NP derived from nuclear cell extract, which included all the milieu of host proteins present. What went unidentified at the time were the host factors that were required. This came later, when a yeast two-hybrid screen, which employed the LexA/B42 system with a LacZ LUE2 dual reporter using a fusion of the NP protein from A/PR/8/33 IAV as bait and HeLa cell cDNA library as prey. One of the most notable proteins identified in this screen was UAP56, originally called several different names including BAT-1, NPI-5, and RAF-2p48[187].

This screen also identified the minichromosome maintenance complex mentioned earlier, as well as karyopherin alpha as hits[188,189]. While a yeast two-hybrid screen is a good hypothesis generation tool, it requires other validation as it is prone to artefacts. However, the authors noted this protein was independently identified by blindly looking at the stimulatory ability of fractions of HeLa nuclear cell extract. Increasingly fine fractionation narrowed down individual protein factors responsible for IAV replication synthesis[173]. Proteolysis followed by mass spec identified UAP56 this way orthogonally to their yeast two-hybrid screen.

Curiously, UAP56 which was cloned and recombinantly expressed had significantly lower stimulatory activity than that which was isolated from source. The authors note that another protein designated RAF-2p36 co-purified with UAP56 from HeLa cell nuclear extract fractionation. They propose that it may be a cofactor or regulator of UAP56, or even have stimulatory properties in its own right. This possible activator of UAP56 not present in their recombinant samples might account for this difference. Since it was never identified, it is possible that it might have been the THO subunit hprt1.

Several key findings were established by this paper about the relationship between NP and UAP56. Firstly, GST-tagged recombinant UAP56 will only co-precipitate NP from vRNPs treated with micrococcal nuclease, which degrades all RNA, but not with intact vRNPs. This established the theory that UAP56 only interacts with RNA-free NP. Secondly, UAP56 increases the synthesis of vRNA in a dose dependent manner. Of note however, this effect was only significant for model RNA, and not actual viral transcripts, and the enhancement effect was inversely proportional to the transcript length. Third, coprecipitation experiments were carried out of GST fusions to truncations of UAP56 with full length NP, or full length UAP56 with GST fusions to truncations of NP. From these, it was claimed that the N-terminal 20 residues of NP were sufficient to interact with UAP56 and that UAP56 a.a. 249-428 (approximately RecA-C) is sufficient to pull down NP. Finally, it was demonstrated that UAP56 increases the fraction of NP bound to RNA in solution. The findings of subsequent papers, as well as evidence presented in this dissertation has corroborated most of these findings. The findings about the minimal binding elements of UAP56 and NP are contentious in light of more recent data and crystal structures not available when this was published. This will be elaborated in a later section.

These authors later expanded on the NP-UAP56 interaction. They demonstrated through an in vitro replication assay that addition of recombinant UAP56 and NP to isolated vRNPs led to substantially longer cRNA on the order of hundreds of nucleotides in length in comparison to replication with NP alone, whose cRNAs were mostly less than 20 nt. Additionally, Ni affinity pulldown of the NP used in the assay revealed its incorporation was selectively enriched in the longer cRNAs with UAP56 present, but selectively enriched in very short cRNAs in its absence. The authors point to this as a role for UAP56 in the facilitation of NP loading onto cRNA, and indirectly as a result of the more efficient NP incorporation, as a processivity factor for polymerase (Fig. 1.7). Importantly, these in vitro experiments were also carried out in the presence of the non-hydrolysable nucleotide analogue ATP-γ-S. From this, the authors contend that the action of UAP56 in viral replication is independent of its ATPase and helicase activity. This later paper also provided the first evidence in cells for a UAP56 role in IAV infection. siRNA mediated knockdown of UAP56 reduced levels of both cRNA and vRNA in HeLa cells[179].

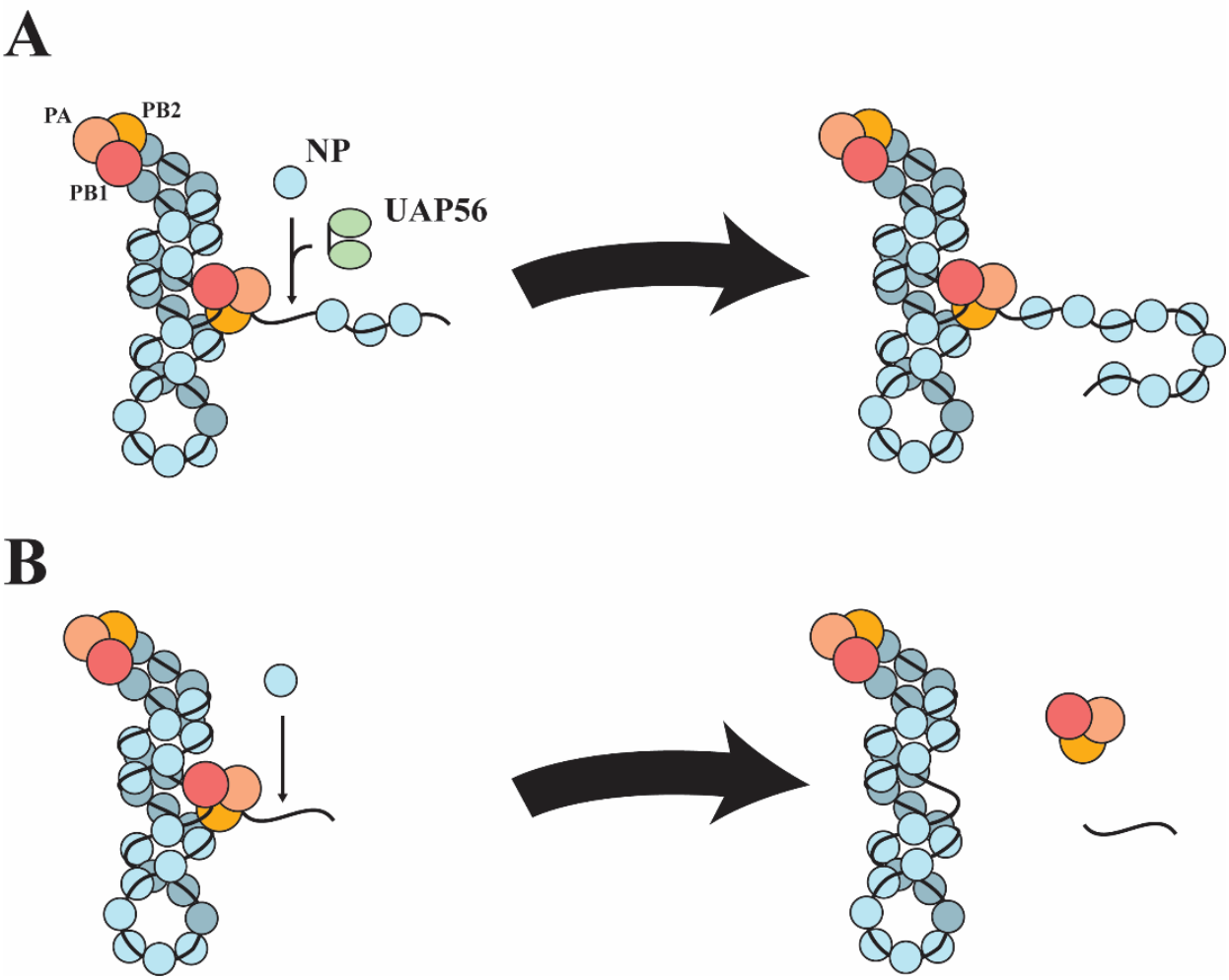

Figure 1.12 – Influenza Viral Replication

A) Viral replication and the role of UAP56 in recruiting NP to nascent vRNA and cRNA during elongation to prevent abortive replication. **B)** Abortive replication as it occurs without UAP56 or other host factors

The most recent paper on the interaction of UAP56 with NP supports the conclusions of the first two in that it provides further biophysical evidence of interaction, while recapitulating in vitro experiments with the NP-RNA complex. However, its novel contention is that UAP56 interacts preferably with a trimeric form of NP. Evidence for this was shown by gel filtration whereby the UAP56-NP complex migrated with the approximate mass of six NP plus two UAP56 in

complex. The distribution of UAP56 in the size exclusion fractions was shifted toward earlier elution with wild type NP. By contrast, with the monomeric NP mutant R416A, UAP56 ran mono- dispersedly as a 100 KDa peak. Analysis with atomic force microscopy showed a singular gaussian distribution centered on approximately 45 nm as the length of the major axis of the bi-lobed particles observed, the authors put these data together into a model in which UAP56 interacts with NP as a dimer with each monomer of UAP56 interacting separately with an NP trimer, and that this is the physiological basis by which UAP56 recruits NP to vRNA[190]. This point is one area in which the findings of this dissertation challenge. The problems with this model, and the alternative model put forth by this dissertation are discussed in a later section. It is with this background that we entered into the project and our principle goal was to refine this existing model by working to obtain atomic detail that was lacking about the precise interface of the UAP56 NP interaction.

CHAPTER 2

IDENTIFICATION AND CHARACTERIZATION OF UAP56 BINDING ELEMENTS TOWARDS NP

2.1 – Introduction

As previously mentioned, there have been several papers which have described the UAP56 NP interaction[173,179,187,190]. When it was originally identified in a yeast two-hybrid screen, it was claimed that the minimal interacting region of UAP56 was the C-terminal RecA domain[187]. This was built on by a study which developed a model that emphasized the functional importance of NP in trimeric form[190]. Both of these were conclusions from other groups that we sought to test so as to find the optimal binding conditions. This chapter describes the initial part of our studies. It involves the biochemical characterization of the general aspects of the UAP56 NP interaction. This began with the purification of NP and UAP56, and the testing of their ability to bind. This was in order to reproduce previous findings in the literature and provide a launch point for examining the interaction in greater detail using truncation constructs of UAP56. From there we examined how binding was influenced by a variety of factors. Specifically, NP strain and oligomeric state were examined for their effect, and the effect of ionic strength was documented. The results of these experiments refined our understanding of several key points of knowledge about this interaction. The driving purpose for our early studies was not only to provide a better understanding of the mechanics of this interaction and the conditions under which they occur in the context of the cellular environment. It was also in order to pave the way for more refined

structural studies to try to capture atomic detail about the complex between UAP56 and NP, which was the ultimate goal of the study.

2.2 – Methods

2.2.1 – Molecular Cloning of NP

The codon-optimized coding sequence for the influenza A/Puerto Rico/8/1934 (H1N1) NP gene was purchased from Sino Biological Inc. The full-length NP gene was PCR amplified, digested with NcoI and NotI with ligated into the NcoI and NotI sites of the pET28a vector to generate NP with a C-terminal His tag. The R416A mutation for NP* and WSN NP R416A was performed by overlap extension PCR methods.

2.2.2 – Molecular Cloning of UAP56

The PCR-amplified intron-less sequence for human UAP56 was digested with BamHI and NotI. It was then ligated into a pGEX4T1-based vector which was digested with BamHI and NotI and treated with alkaline phosphatase. The vector was modified so that the N-terminal GST tag contained a spacer with the TEV cleavage site between the tag and the coding sequence. Truncations of UAP56 were made by overlap extension PCR with primers designed to partially overlap the desired boundaries introducing a BamHI cut site upstream of the starting residue, and a stop codon followed by a NotI cut site at the desired end of the sequence. This was followed by digestion and ligation back into the a pGEX4T1-based empty plasmid backbone. The above

method was also used in the creation of GFP-UAP56 constructs, except utilizing a pProEX HTb vector.

2.2.3 – Protein Expression

For all protein expression, plasmids were transformed into chemically competent E. coli Rosetta cells under selection of either ampicillin for UAP56 constructs or kanamycin for NP. Cells were grown in LB at 37 °C to OD600 of 0.6 – 0.8, then induced with 0.5 mM IPTG for 16 hours at 16 °C.

2.2.4 – NP Purification

For NP proteins, cells were resuspended in lysis buffer A (50 mM Tris HCl pH 8.0, 500 mM NaCl, 20 mM imidazole, 0.5 mM TCEP, 1 μg/ml aprotinin, 1 μg/ml pepstatin, 1 μg/ml leupeptin, and 0.1 mM AEBSF) and lysed by homogenizer. Clarified lysate was bound to Ni-sepharose resin in batch mode for 1.5 hr. Resin was then washed extensively with lysis buffer A followed by high salt buffer (50 mM Tris HCl pH 8.0, 1.5 M NaCl, 0.5 mM TCEP, 1 μg/ml aprotinin, 1 μg/ml pepstatin-A, 1 μg/ml leupeptin, and 0.1 mM AEBSF). Resin was then resuspended in lysis buffer A and RNase A was added to 90 μg/ml and allowed to digest overnight. This was followed by extensive washing with lysis buffer A and elution with 50 mM Tris HCl pH 8.0, 300 mM NaCl, 250 mM imidazole, 10% v/v glycerol, 0.5 mM TCEP, 1 μg/ml aprotinin, 1 μg/ml pepstatin-A, 1 μg/ml leupeptin, and 0.1 mM AEBSF. Protein was then loaded onto a 1 ml Heparin HP column and eluted in 10 mM HEPES pH 7.0, 0.5 mM TCEP with a linear gradient of NaCl. 0.1 mM AEBSF was added to fractions with protein before pooling the purest fraction

as judged by running SDS-PAGE. The protein was further purified using a Superdex 200 10/300 GL column equilibrated in 10 mM HEPES pH 7.0, 150 mM NaCl, and 0.5 mM TCEP. Purified proteins were flash frozen and stored at -80 °C. Protein concentration was estimated based on A280. In all cases A260/A280 was less than 0.6, indicating that purified NP protein was free of RNA. All purification steps were performed on ice or in a cold cabinet.

2.2.5 – UAP56 purification

For GST tagged UAP56 proteins, cells were lysed in lysis buffer B (50 mM Tris HCl pH 8.0, 500 mM NaCl, 0.5 mM TCEP, 5 µg/ml aprotinin, 2 µg/ml pepstatin-A, 2 µg/ml leupeptin, and 0.5 mM AEBSF). Clarified lysate was bound to glutathione sepharose 4B resin for 30 minutes. The resin was washed extensively with lysis buffer B and eluted with 50 mM Tris pH 8.0, 500 mM NaCl, 20 mM glutathione, 0.5 mM TCEP, 10 % v/v glycerol 2 µg/ml aprotinin, 2 µg/ml pepstatin-A, 2 µg/ml leupeptin, and 0.2 mM AEBSF. In some preparations the GST tag was cleaved by addition to the resin of several mg of TEV protease with a non-cleavable GST tag for 16 hours, rather than elution with glutathione. In this instance, the tag-less UAP56 passed in the flowthrough. Protein product was then passed over a HiTrap Q FF column and eluted in 10 mM Tris pH 8.0, 0.5 mM TCEP with a linear gradient of NaCl. 0.1 mM AEBSF was added to fractions with protein before pooling. Protein was further purified over a Superdex 200 10/300 GL column in 10 mM Tris pH 8.0, 150 mM NaCl, and 0.5 mM TCEP. Protein fractions from each step were assessed by PAGE and pooled very stringently, and concentration calculated by A280. Purified proteins were flash frozen and stored at -80 °C.

Purification of GFP constructs was identical except that the resin used in the affinity chromatography step was Ni-sepharose resin and the elution buffer contained 250 mM imidazole rather than glutathione. Synthesized FITC-UAP56-NTE$_{1-19}$ was ordered from GenScript.

2.2.6 – GST UAP56 co-precipitation of NP

GST tagged UAP56-FL, UAP56-Core, UAP56-NTE$_{1-38}$, GST-UAP56-NTE$_{1-19,}$ or GST alone were added to binding buffer (10 mM Tris HCl pH 8.0, 100 mM NaCl, 0.5 mM TCEP, and 0.1% IGEPAL CA-630) at 2 μM in the presence or absence of 2 μM of NP. Binding to glutathione sepharose 4B resin was allowed to proceed for 30 min at room temperature. Resin was washed three times with binding buffer, and bound proteins were eluted and analyzed by Coomassie-stained SDS-PAGE gels. Pulldowns with NP* followed the same protocol but with binding buffer containing 50 mM NaCl. Pulldowns examining the salt sensitivity were performed in the same way except 5 μM of GST-UAP56-FL and 5 μM of either NP*, PR8 WT, or WSN NP was used in a buffer of 10 mM Tris HCl pH 8.0, 0.5 mM TCEP, and varying amounts of NaCl.

2.2.7 – Native PAGE Electrophoretic Mobility Shift Assay

Samples were loaded as 10 ul on a 5% native polyacrylamide gel that was prepared with 45 mM Tris and 45 mM boric acid and pre-run in the same buffer. Gels were run at 100V at 4 °C.

All binding titrations were conducted in a buffer containing 10 mM Tris pH 8.0, 150 mM NaCl, 0.5 mM TCEP, 10% v/v glycerol, and allowed in incubate for 30 minutes at room temperature before running. Gels were visualized with Coomassie stain. Native EMSA titrations using GFP-

UAP56 constructs were conducted in the same manner, but with a fixed 200 nM concentration of the GFP-UAP56 protein and increasing amounts of NP*. These gels were visualized by fluorescence with a Typhoon FLA 9000 biomolecular imager using the Alexa Fluor 488 filter with the highest detector gain which produced unsaturated signal.

2.2.8 – Microscale Thermophoresis

10 ul of 50 nM FITC-UAP56-NTE$_{1-19}$ were added to 10 ul from each of two-fold serial dilutions of WSN R416A NP in a buffer of 150 mM NaCl 10 mM HEPES pH 7.0, 0.5 mM TCEP. This was mixed and incubated at room temperature for 10 minutes. Each reaction was then transferred to completely fill a regular MST capillary. Capillaries were then pre-scanned on a Monolith NT.115 to ensure no adsorption had occurred, and thermophoresis measured with the fluorescence excitation set to 20% on the blue filter and the infrared set to medium with 5s pre-scan time, 25s on time and 5s recovery time for each trace. The response curve was calculated based on the difference in normalized fluorescence at 5s into the IR pulse compared to 1s pre-pulse. Data was included from two separate replicates and fit to the standard Hill Equation using a custom R-script.

2.2.9 – Fluorescence Anisotropy

NP* in 10 mM HEPES pH 7.0, 150 mM NaCl, 0.5 mM TCEP was serially diluted by a factor of 4/3 21 times beginning at 8 μM. In a 394-well black fluorescence plate 12 ul of each serial dilution or buffer for blank were added to each well in one row and 12 ul of 80 μM FITC-UAP56-NTE$_{1-19}$ was added and briefly mixed. Anisotropy was measured using a polarization

filter on a Take3 automated plate reader with a 485 nm excitation maximum and 528 nm emission maximum with gain set to 65 for both parallel and perpendicular detectors. Anisotropy was automatically calculated by the software and blank corrected. The resulting response curve was fit to the standard Hill Equation using a custom R-script.

2.3 – Results

2.3.1.1 – Purification of PR8 NP

While UAP56 had been previously expressed and purified in the lab, NP had not. Our expression and purification strategy was thorough, to produce the purest protein possible with the anticipation that the material may be used for crystallography. We chose a C-terminal His tag as this had been used in published crystal structures of NP[153]. Growth in E. coli with LB media led to high expression and an abundance of starting material.

Heparin was used as a second step after the nickel affinity and RNase treatment. Because of the high positive charge of NP (+13 at pH 7.0), hard cation exchange was quite effective in removing protein which non-specifically bound during the Ni-Affinity step, with NP only eluting at ~950 mM NaCl (**Fig 2.1 B**). The small pre-peak in heparin exchange may represent a small fraction of NP in monomeric form, however, it was not used in the final SEC step.

The final purification step of SEC produced highly pure protein (**Fig. 2.1 A** and **C**). Of note, while there was one main peak, the protein was not mono-dispersed. A significant shoulder and a trailing tail on opposite sides of the central peak at around 12 ml likely indicate a variety of multimeric states in equilibrium. This is as expected from the literature, as the trimeric form of

NP is thought to be the predominating form in vitro, and it has been proposed to have a physiological significance[190], although this idea is disputed[191]. The elution volume in our chromatograms corresponds to a mix of mostly trimer and dimer.

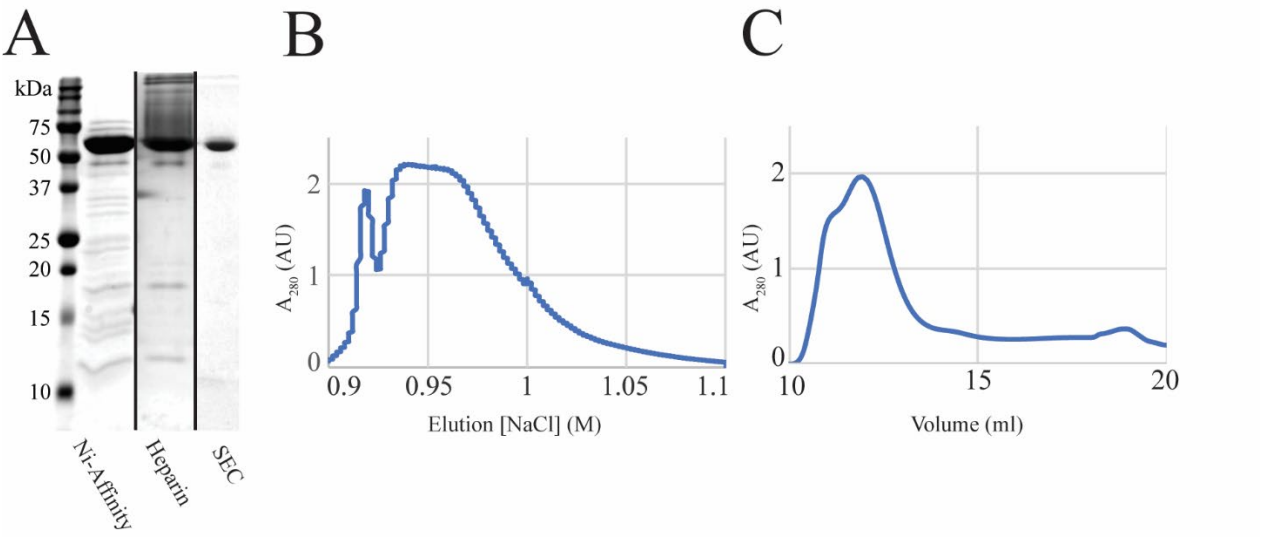

Figure 2.1 – Purification of PR8 NP

A) Coomassie stained samples of PR8 NP at various purification steps on SDS-PAGE **B)** Heparin ion exchange chromatogram plotted based on eluting salt **C)** Size exclusion chromatogram

In the first iteration of its purification, NP had not been treated with RNase A. Ideally, this purification protocol would have been used throughout our study because downstream uses of NP require RNA, to which even a tiny amount of RNase contamination would cause problems. We tried to remove all bound nucleotide simply with a high salt wash step during nickel affinity pulldown and passage over a heparin cation exchange column. The final SEC step consistently showed that this was insufficient to obtain RNA-free NP as the A260/A280 was always too high.

For this reason, the overnight on-resin RNase A incubation was added, as this had been used by others during NP purification[153,154,190,192]. Despite our initial misgivings, we did not experience any problems with RNA degradation in downstream assays.

2.3.1.2 – Purification of UAP56

The purification of UAP56 followed standard protocols previously used in the lab. For the GST-tagged constructs, initial pulldown over glutathione sepharose gave good separation from other proteins. However, degradation of UAP56 presented a problem (**Fig. 2.2**). High concentrations of a multi-factor protease inhibitor cocktail were used in lysis as well as added immediately to all chromatographic fractions except the final SEC. Lysate processing and glutathione resin incubation times were minimized, and ion exchange fractions and SEC fractions were cut very stringently.

Even with all of these steps, we could not completely eliminate degradation. Much of this problem is due to the nature of the protein. UAP56 has a flexible, loosely structured N-terminus with a glycine-serine spacer and TEV cleavage site added on by the GST tag. In addition, there is a flexible linker between the two RecA domains. The protein has a lot of protease-sensitive sequences that make it particularly vulnerable to degradation, despite our efforts. Constructs of UAP56 truncations which lacked these regions were consequently purified with much less degradation. Similar problems with degradation were encountered for GFP tagged constructs of UAP56 in which GFP was also connected to the N-terminus by a serine-glycine linker. Despite

this, the final purified product can conservatively be estimated to be at least 90% pure in all cases.

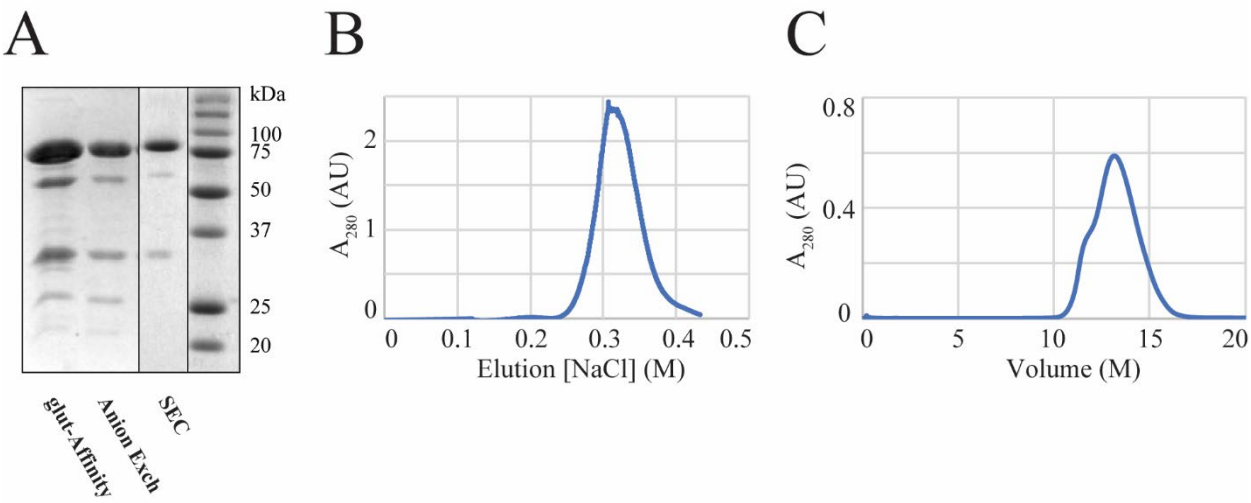

Figure 2.2 – Purification of GST-UAP56

A) Coomassie stained samples of GST-UAP56 at various purification steps run on SDS-PAGE. **B)** Heparin ion exchange chromatogram plotted based on eluting salt **C)** Size exclusion chromatogram

2.3.1.3 – UAP56 can Co-Precipitate NP

Having purified both components of the system, we started by testing the most basic interaction to confirm the literature reports that UAP56 is capable of binding NP. The simplest way to way to confirm this is by pulldown of Full length UAP56 (GST-UAP56-FL) over glutathione sepharose resin by its N-terminal GST tag and observing whether NP is pulled down and co-eluted by glutathione. As shown in (**Fig. 2.3**), bands of both GST-UAP56-FL and NP can clearly be seen in the glutathione eluate when the two are incubated together.

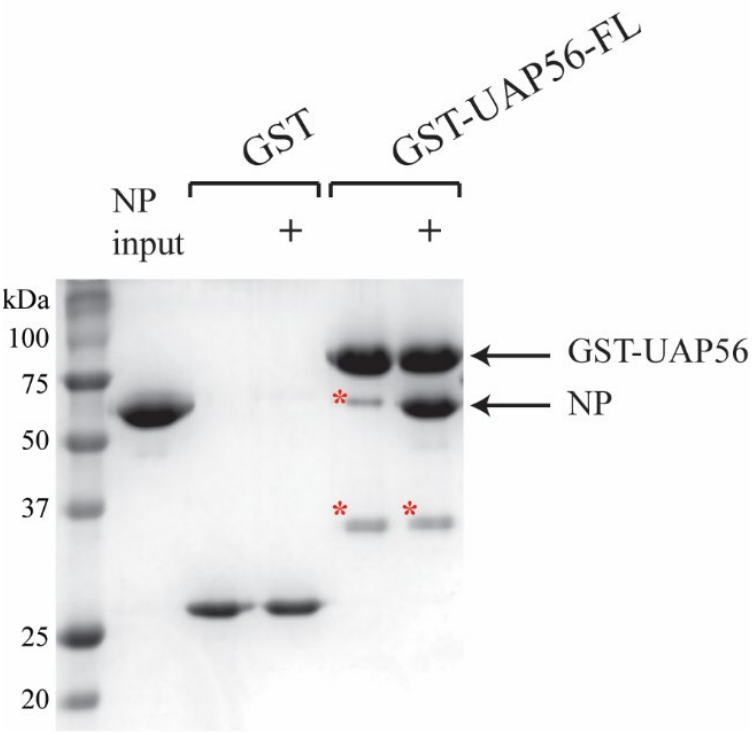

Figure 2.3 – PR8 NP Co-precipitates with GST-UAP56

NP Input and glutathione eluate fractions for either GST tag or GST-UAP56-FL with NP present or absent from the mix. Run on SDS-PAGE stained with Coomassie. Red asterisks indicate UAP56 degradation fragments. Figure modified from Morris *et al*, 2020[3].

The fact that NP cannot be co-precipitated by GST alone shows that the binding is not an artefact. From this simple experiment we were able to reproduce findings by other groups and confirm that, at least under a certain set of experimental conditions, binding can be seen between UAP56 and NP in vitro.

2.3.1.4 – The Purification of Monomeric NP

As mentioned, NP exists in an ensemble of multimeric states in equilibrium with each other and with the monomeric form. The multimerization state of NP is known to depend on both concentration and the ionic strength of the buffer[155], and the kinetics between oligomerization states was shown to be quite slow[153,155]. Furthermore, a key finding from a previous report on UAP56 NP interaction was that UAP56 binds highly preferentially to RNA-free NP in a trimeric state[190]. If this is indeed true, then for any given binding assay, the result depends on the multimerization state of NP. This presents a challenge in interpreting results because it is very difficult to know the population of the different multimerization states when performing any in vitro assay across varying buffer conditions. How close to equilibrium these states are is also difficult to estimate as the transition between states is very slow. To eliminate this uncontrolled variable from binding assays and also to test if a multimeric form of NP is required for UAP56 interaction, we generated a monomeric mutant.

Our monomeric form of NP was created by mutating a critical arginine within the flexible tail loop of NP that inserts into the neighboring monomer and makes a salt bridge with glutamate 339 of its neighbor. It had been previously well established in the literature that this mutant behaves as a monomer in solution[153–155]. We generated the R416A mutant from our PR8 NP, which will be referred to hereafter as NP*. NP* was expressed and purified in the same manner as NP, with Ni-affinity pulldown and on-column RNase treatment. While the steps in purification were the same, NP* eluted at much lower salt over the heparin ion exchange column with a main peak around 600 mM NaCl (**Fig. 2.4 C**).

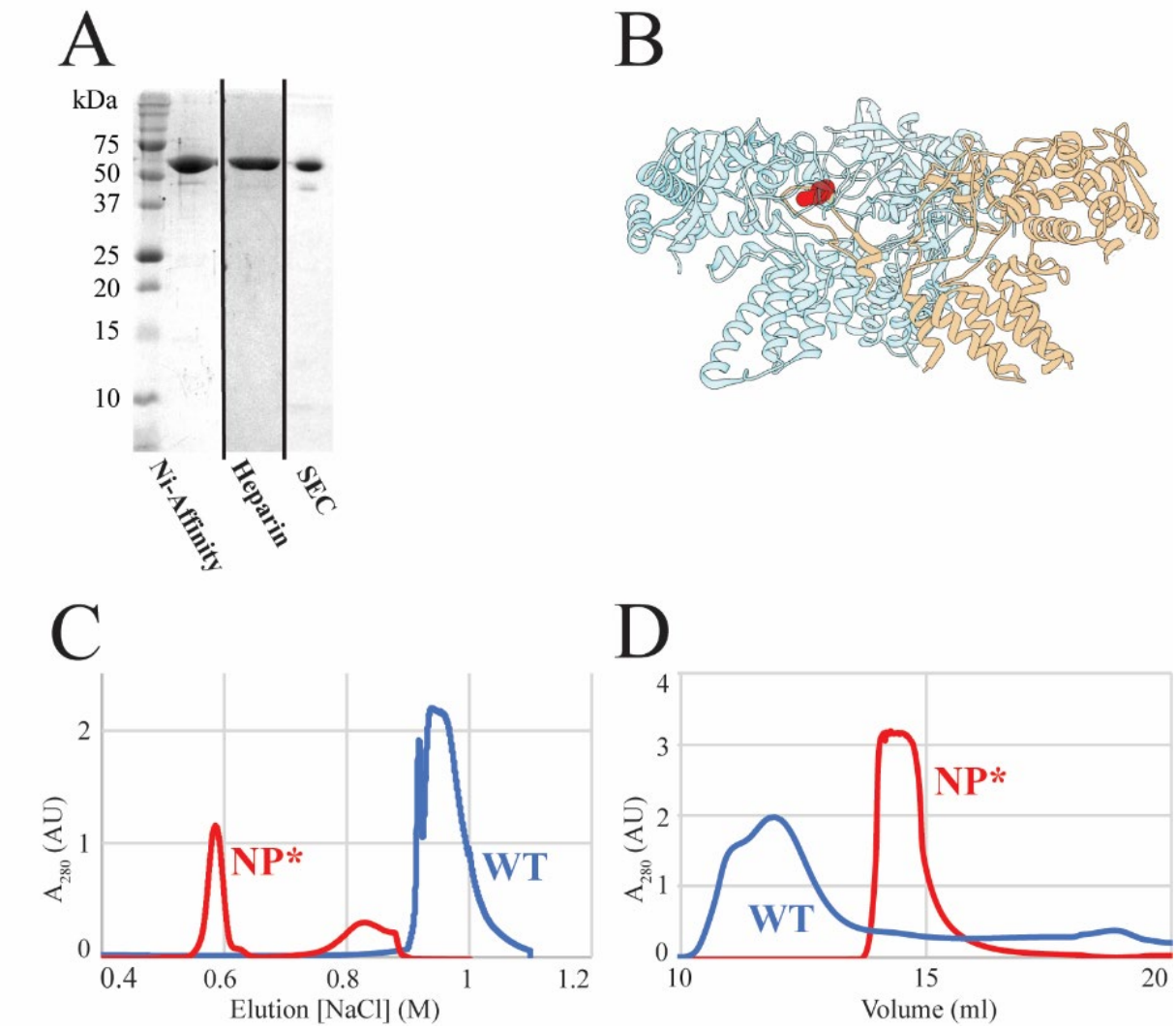

Figure 2.4 – Purification of NP*

A) Coomassie stained samples of PR8 NP at various purification steps on SDS-PAGE **B)** A ribbon diagram of NP A/WSN/1933, from PDB: 2IQH[153]. Two subunits are transparent and R416A is highlighted to demonstrate how the tail loop inserts. Made using Chimera[69] **C)** Heparin ion exchange chromatogram plotted based on eluting salt comparing PR8 WT and NP* **D)** Size exclusion chromatogram comparing WT to NP*

There was a second peak at about 850 mM NaCl. Though this peak was identified as NP by SDS-PAGE, it was not included subsequent steps. NP* is known to have considerably weaker

binding to RNA compared to the WT[155]. As heparin and RNA have similar physical properties, this likely explains the difference in elution over the heparin column for NP* as compared to WT. When run over SEC, NP* ran as a single sharp peak eluting at about 14.5 ml off of the Superdex 200 column, considerably later than for WT (**Fig. 2.4 D**). This indicates that even at the high concentrations at which it was injected over the column, NP* is well behaved and uniformly occupies a monomeric state in solution.

2.3.1.5 – UAP56 can Co-Precipitate NP*

Having purified NP*, we sought to examine whether previous claims that UAP56 only interacts with trimeric NP had merit[190]. As was done for the NP WT, NP* was incubated with GST-UAP56-FL over glutathione sepharose resin and was observed to be co-precipitated in the eluate (**Fig. 2.5**). While binding was not as robust as for NP WT, we were able to show that this interaction can still occur with monomeric forms of NP under lower salt conditions which are notably gentler than those used in previous reports[190].

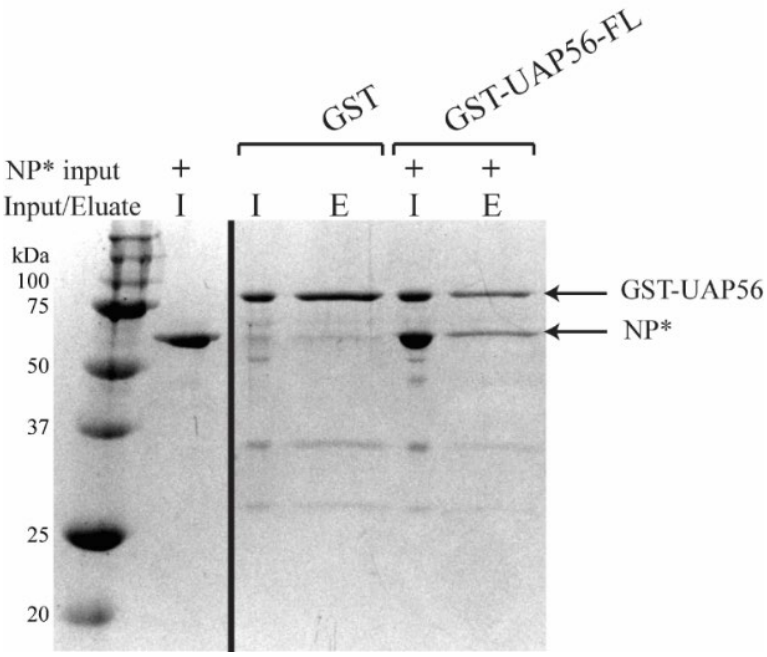

Figure 2.5 – NP* Co-precipitates with GST-UAP56

NP* input and input plus glutathione eluate fractions for either GST-UAP56-FL with NP present or absent from the mix, Coomassie stained SDS-PAGE. Arrows indicate relevant proteins, other bands are UAP56 degradation fragments.

2.3.2 – Characterization of A/WSN/33 NP

2.3.2.1 – Purification of WSN NP

While we were investigating the role that the oligomeric state of NP had on its interaction with UAP56, we were also curious about the strain specificity of this interaction. Many of the interactions observed between viral and host proteins are strain specific and often account for the host range of the virus. The PB2 protein in avian H5N1 strains, for example, is only capable of interacting with avian host factors, but not mammalian ones[193–195]. All studies done up to this point on the UAP56-NP interaction had used the common laboratory strain PR8. To test whether this interaction was specific to PR8 or a broader mechanism of influenza exploitation of host

factors we expressed and purified A/WSN/33 NP in the same manner as was done for PR8. Other than a slightly earlier elution over heparin (**Fig 2.6 B**) and a more symmetrical SEC profile (**Fig 2.6 C**), WSN behaved almost identically to PR8 WT, including the presence of a small monomer population.

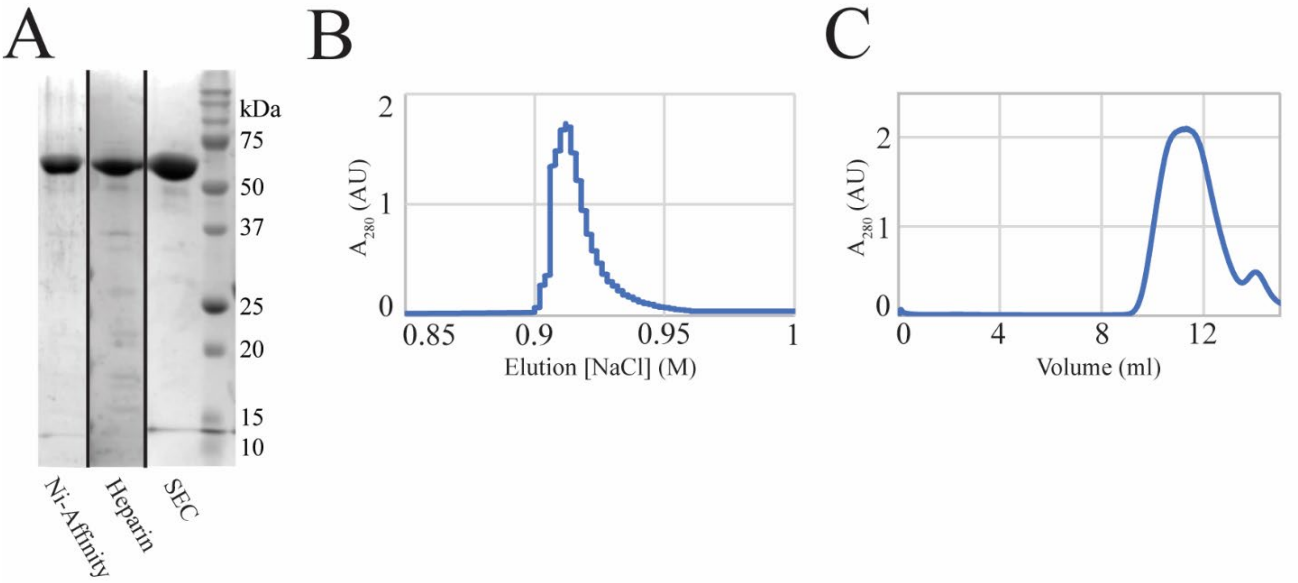

Figure 2.6 – Purification of WSN NP

A) Coomassie stained samples of WSN NP at various purification steps run on SDS-PAGE **B)** Heparin ion exchange chromatogram plotted based on eluting salt **C)** Size exclusion chromatogram

2.3.2.2 – UAP56 can Co-Precipitate WSN NP

Testing the binding of WSN NP to GST-UAP56-FL was performed in the same way as was done for PR8. GST-UAP56-FL was precipitated over glutathione sepharose resin and when WSN NP was added it was observed to coprecipitate with GST-UAP56-FL (**Fig 2.7**)**.** This result

demonstrates that interaction with UAP56 is not a feature unique to the PR8 strain of IAV, and that it can be seen for other strains as well. This does not imply that it ubiquitous, however.

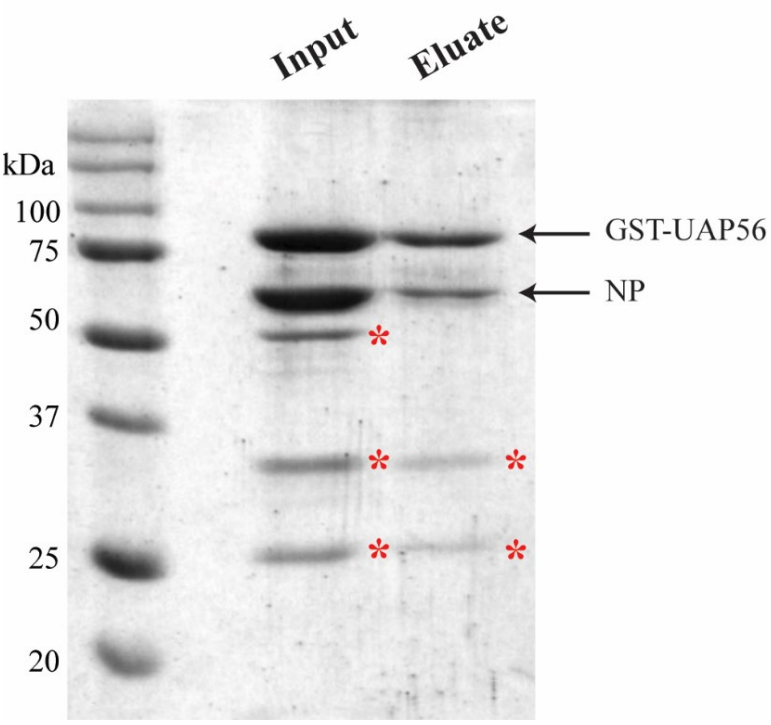

Figure 2.7 – WSN NP Co-precipitates with GST-UAP56

Input and glutathione eluate fractions for GST-UAP56-FL and WSN NP mixture. Run on SDS-PAGE stained with Coomassie. Red asterisks indicate UAP56 degradation fragments. NP incubated alone with glutathione resin did not bind (data not shown).

2.3.3 – UAP56-NTE is the primary binding site for NP

2.3.3.1 – Construction of UAP56 Truncation Mutants

To dissect the minimal region of UAP56 which could bind to NP, the lab had made N-terminal GST fusions to truncations of UAP56. A fusion of the N-terminal extension (hereafter referred to

as NTE) consisting of the mostly unstructured residues was made, as well as a fusion consisting of all residues outside the NTE (a.a. 44- 428). All the constructs containing only residues 44-428 (the RecA domains) will hereafter be referred to as UAP56-Core. For some applications, the GST tag was removed from UAP56-Core by TEV protease. Additionally, fusion constructs were

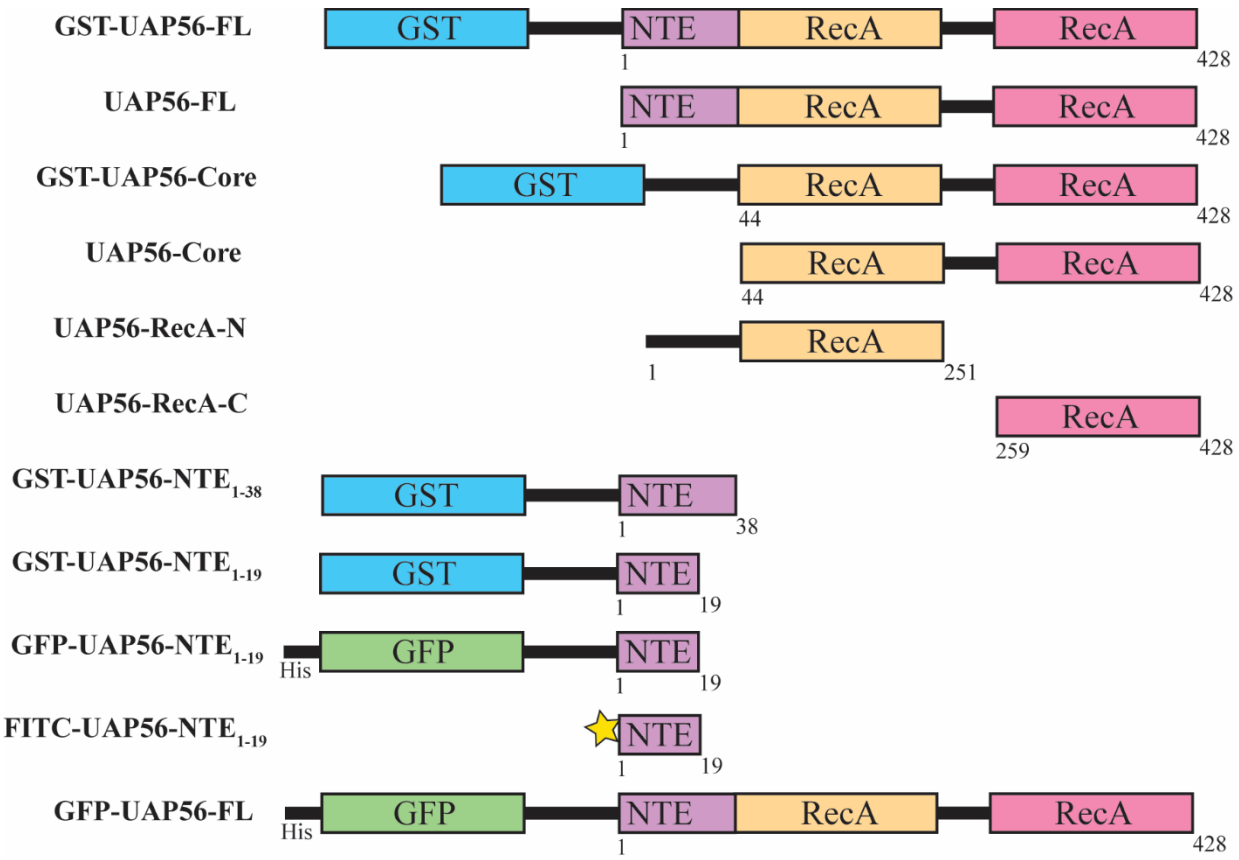

Figure 2.8 – Constructs of UAP56 Utilized

A domain map of the constructs of UAP56. Constructs which lack a tag were produced as GST fusions followed by TEV cleavage. FITC-UAP56-NTE$_{1-19}$ is a synthesized peptide with a FITC fluorophore on the N-terminal amine.

made consisting of the NTE with the N-terminal RecA domain (a.a. 1-251, UAP56-RecA-N) or of the C-terminal RecA domain alone (a.a. 259-428, UAP56-RecA-C). For these, the GST tag was used for affinity purification only and was removed by on-column TEV digestion. A complete list of all truncations of UAP56 used in our study and their nomenclature can be seen in (**Fig 2.8**).

2.3.3.2 – Native EMSA Reveals Interaction between NP and UAP56 N-terminus

The least stringent method to look at protein-protein interactions is by native gel electrophoretic mobility shift assay (native EMSA). At pH 8.0, NP is highly positively charged, and will run towards the cathode in an electrolytic cell, consequently it will barely enter the gel. UAP56 is an acidic protein and will migrate in the opposite direction towards the anode and be pulled into the gel. The formation of a complex between these leads to a change in the migration of UAP56 whereby its movement into the gel is restricted and the band of UAP56 protein depletes while a band corresponding to the complex grows stronger with increasing amounts of NP. Depending on the construct, this band may not enter the gel at all and only depletion of the UAP56 band is seen. As a positive control, a 5 μM equimolar mixture of UAP56 and NP* led to the depletion of the band corresponding to unbound UAP56 when compared to the same amount of UAP56 run alone. This is consistent with our coprecipitation experiment done with NP*. What was novel however, was that we observed that depletion of the free UAP56 band could be seen when NP* was incubated with a GST fusion of UAP56 residues 1-38 of the NTE (GST-UAP56-NTE$_{1-38}$). The RecA domains outside the NTE (UAP56-Core) did not produce a gel shift at 5 μM

concentration with NP* (**Fig 2.9 A**). This was the first indication we had that the NTE is responsible for NP binding to UAP56.

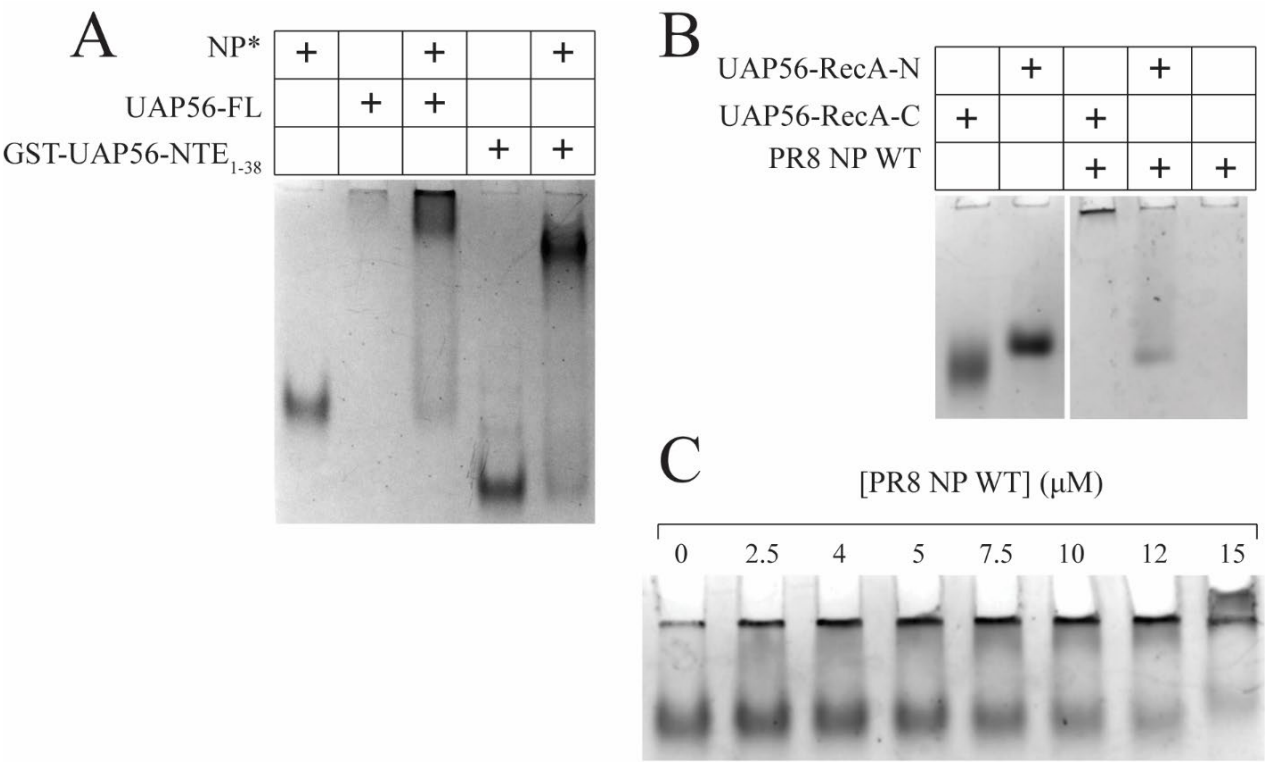

Figure 2.9 – Native EMSAs Shows NP interaction with UAP56-NTE and RecA Domains

A) Native EMSA with 5 μM of the indicated protein in isolation or mixed. Note that NP* barely enters the gel due to its high positive charge. **B)** Native EMSA with 15 μM of the indicated protein in isolation or mixed. **C)** A native EMSA titration with a fixed 5 μM of UAP56-Core and increasing concentrations of PR8 NP WT. Note that difference in the migration for the 15 μM sample is due to a defect in the well.

The paper which originally identified the UAP56 NP interaction found RecA-C to be the minimal component for NP interaction[187]. While UAP56-Core was not seen to interact with NP* at 5 μM, PR8 WT NP, as used in that earlier paper could be seen to interact with UAP56-core at

higher concentrations in a titration of NP up to 15 μM (**Fig. 2.9 C**). The individual RecA domains, UAP56-RecA-N and UAP56-RecA-C were also seen to interact with PR8 NP WT at an equimolar concentration of 15 μM. Together these native EMSAs suggested that the most important region of UAP56 for interaction with NP is the NTE, but also suggest some weaker interaction can be seen from the RecA domains either in isolation or as UAP56-Core. It is worth noting however, that native EMSA as a method is prone to artefacts and can also pick up non-specific interactions through a molecular caging effect. Because of the limitations of native EMSA as a technique, we went back and looked at the interaction of UAP56 fragments with NP and NP* by coprecipitation.

2.3.3.3 – UAP56-NTE can Co-Precipitate NP and NP*

In the same manner as done for the first experiments, we looked at the ability of GST fusions of UAP56 truncations to coprecipitate PR8 NP WT over glutathione sepharose resin. At 2 μM equimolar concentrations of several UAP56 GST fusions with PR8 NP WT, we observed as a positive control that GST-UAP56-FL could coprecipitate NP (**Fig 2.10 A**). We further observed, in agreement with our native EMSAs, that GST-UAP56-NTE$_{1-38}$ was sufficient to coprecipitate NP with approximately the same efficiency as for GST-UAP56-FL, as judged qualitatively by the intensity of the NP band in the eluates. In further agreement with our native EMSA data, GST-UAP56-Core was also capable of coprecipitating NP. This coprecipitation was less efficient than that seen for either GST-UAP56-FL or GST-UAP56-NTE$_{1-38}$, which agrees with our finding of weaker interaction between RecA domains and NP than for NTE. We were further able to show that the strong interaction seen with GST-UAP56-NTE$_{1-38}$ held true when NP* was used

instead of PR8 NP WT (**Fig. 2.10 B**). As with the native EMSA experiments, no interaction was seen when NP* was tested with GST-UAP56-Core.

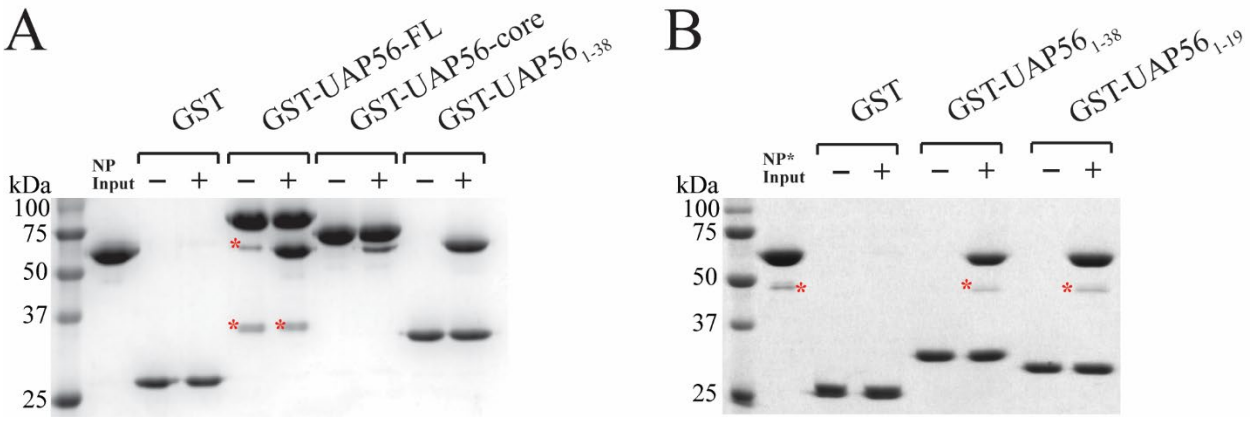

Figure 2.10 – Coprecipitation Demonstrates UAP56-NTE is Sufficient for Binding

A) NP Input and glutathione eluate fractions for either GST tag, GST-UAP56-FL, GST-UAP56-Core, or GST-UAP56$_{1-38}$ with NP present or absent from the mix. Run on SDS-PAGE stained with Coomassie. Red asterisks indicate degradation fragments. **B)** The same experiment as in A) but utilizing NP* with GST-UAP56$_{1-38}$ or GST-UAP56$_{1-19}$. Binding was carried out at 50 mM NaCl rather than 150 mM. Figure modified from Morris et al, 2020[3].

Within the first 38 residues that make up the NTE, we hypothesized that the minimal binding region could be narrowed down further. The latter half of the NTE is poorly conserved between metazoan UAP56 and its yeast homolog Sub2. Within the 19 residues though, is a conserved group of acidic residues common from yeast to metazoans. Given its conservation and the known autoregulatory function of the NTE in controlling ATPase activity[66],we reasoned that it might be important in NP interaction since its high density of positive charge might complement the high density of negative charge on NP. Indeed, when we tested the GST fusion construct with only the

first 19 residues, we saw that GST-UAP56-NTE$_{1-19}$ had essentially the same strength of the interaction, which can be seen for both NP WT and NP*, and that UAP56-Core also contributes to this interaction more weakly, which can be seen only for NP WT.

2.3.3.4 – The UAP56-NTE NP Interaction is Salt Sensitive

One important point about these coprecipitation experiments is that they are sensitive to the ionic strength of the buffer in which they are carried out. Notably the coprecipitations of NP* in **Fig. 2.5** and **Fig 2.10 B** were carried out with 50 mM NaCl present. The pulldowns of WSN NP and PR8 NP WT, by contrast, had 150 mM NaCl. To examine this salt sensitivity further, we performed a coarse titration in which GST-UAP56-NTE$_{1-38}$ was used to coprecipitate either WSN NP, PR8 NP WT, or NP*, each at increasing amounts of NaCl. **Fig. 2.11** shows how PR8 NP WT is robustly pulled down at 150 mM NaCl, while still being capable of binding with reduced strength at 250 mM, while WSN was only capable of binding at 150 mM NaCl, and NP* could only robustly bind at 50 mM NaCl. It is common for protein-protein interactions to be sensitive to salt. Interactions driven by hydrophobic effects tend to become stronger as certain salts are increased within a limited range over low ionic strength. A previous paper had examined the factors affecting NP oligomerization[155]. It found that NaCl- concentrations below 50 mM promoted monomeric NP, whereas above 100 mM the trimeric form dominated. For a protein with very high positive charge at its surface, increasing concentration of NaCl is likely mitigating some electrostatic repulsion between monomers in addition to increasing the solvation penalty for monomers such that it promotes NP multimerization. Increased ionic strength however, disrupts charge-charge driven interactions. The abrogation of binding seen with

increased salt suggests that the NTE-NP interaction is charge-charge driven, which is to be expected. It also suggests that PR8 NP WT has the strongest binding, followed by WSN, and NP* has the weakest. It is to be noted that the physical chemistry of protein solvation is a product of many factors and can be difficult to predict ab initio, and so salt sensitivity is an indirect and qualitative metric of this.

While Na^+ and Cl^- inside mammalian cells are typically low[196], there are many other charged species present with differing effects on solvation that make it hard to extrapolate these simple buffer conditions we used to a physiological context when speculating about either the multimeric state or UAP56 interaction of NP from purely the perspective of physical chemistry. It is known however, that the R416A mutant used for NP* produces an inviable virus and has much reduced binding to RNA. This combined with our observation of higher salt sensitivity of the NP* UAP56 interaction could mean that the observed interaction with NP* does not occur within the cell, or only occurs to a trivial extent. What is important though, is the fact that this interaction can still be seen even with just the UAP56-NTE implies that it is not an inherent and unique property of the NP trimer that is being recognized. The requirement for low ionic strength for NP* binding, which could not occur at 150 mM NaCl, explains why previous reports claimed that the interaction between NP* and UAP56 does not occur, as they used higher salt conditions[190], and is in line with the assertion that NP which is at least capable of multimerizing demonstrates a more robust binding than an obligate monomer.

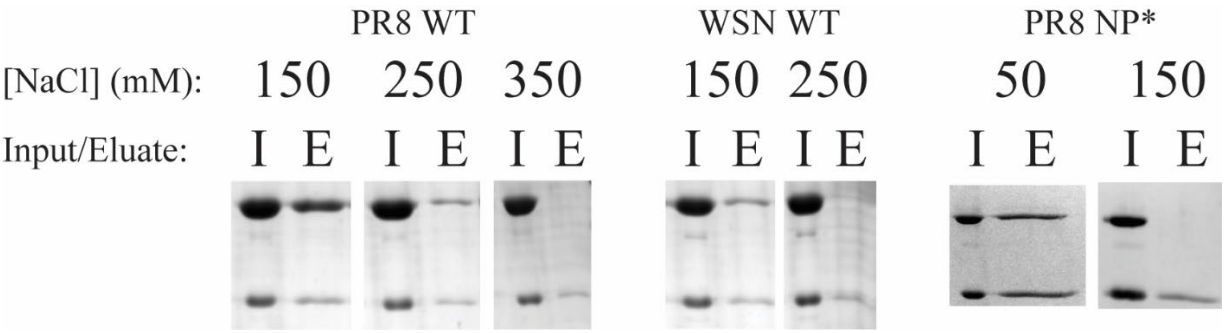

Figure 2.11 – Co-precipitation of NP is Salt-Labile

Input and glutathione eluate fractions of pulldowns over glutathione resin for 5 μM of GST-UAP56-NTE$_{1-38}$ with 15 μM of PR8 WT or WSN NP or 5 μM of GST-UAP56-NTE$_{1-38}$ with 5 μM of NP*, each conducted in the same buffer with varying salt.

2.3.3.5 – Microscale Thermophoresis Verifies UAP56-NTE NP Interaction

Data collected on the UAP56-NP interaction so far had been purely qualitative. In an effort to provide both an orthogonal methodology to validate the interaction that we saw, and as a means to quantify the affinity between the two species, we moved to fluorescence-based techniques. To this end, we engineered, expressed, and purified N-terminal green fluorescent protein (GFP) fusions to either UAP56-NTE$_{1-19}$ or UAP56-FL. Additionally we had UAP56-NTE$_{1-19}$ made as a synthetic peptide with a fluorescein isothiocyanate (FITC) probe chemically conjugated to the N-terminal amine, which we ordered through the company GenScript (see **Fig. 2.8**). Both GFP and FITC tags allow signal to be measured even under very low concentrations, and fluorescence-based techniques are a way to accurately measure dissociation constants in the sub-micromolar range.

The first method we employed is called microscale thermophoresis (MST). It is a relatively new technique which measures the change in diffusibility of a molecular species as a result of binding

to another molecule. One species, in our case FITC-UAP56-NTE$_{1-19}$, is present at a concentration well below the estimated K_D of the interaction between the two species. When a temperature gradient is rapidly applied to a sample in a capillary by an infrared laser, molecules have net diffusion away from the heat source. For a stable gradient, a new equilibrium is reached where the difference between the density of the reporter molecule over the area of focus of the heating laser before and after the infrared pulse is proportional to its diffusibility. For a fluorescently labeled molecule, this thermophoresis is measured as a change in integrated fluorescence intensity over a small area where the temperature gradient is applied (ΔF). Diffusibility depends on several factors, but mostly it is the result of the change in effective mass and size which occurs when the molecule is complexed with another. With the fluorescent species held constant, a titration of the binding partner is added in over a range of several orders of magnitude.

The addition of the binding partner over the low end of the titration range causes little complex formation as both species are well below their dissociation constants. As the K_D is approached, increasing the concentration of the binding partner causes rapid change in occupancy of the fluorescent reporter owing to increased complex formation and consequently changes in ΔF. ΔF then levels off with increasing concentration of the binding partner, as the fractional occupancy of the fluorescent species in a bound state asymptotically approaches one. This forms the classical sigmoidal binding curve seen by other methods when plotted on a log scale[197,198].

When MST was done using FITC-UAP56-NTE$_{1-19}$ as the fluorescent reporter, and WSN NP R416A titrated in, we obtained a binding curve and an estimation of the affinity of their interaction (**Fig. 2.12**). From this we see that even for the monomeric form of NP, the K_D is less than 1 μM. There was an important qualification to this, however. The fit for a binding curve in

which WSN NP R416A and FITC-UAP56-NTE$_{1\text{-}19}$ bind together 1-to-1 with no form of allostery is the Michealis-Menten equation, commonly written as $V = V_{max}[S]/(K_D + [S])$[11]. This famous equation was developed for modeling enzyme catalysis of a substrate at steady-state and assumes that enzyme binding of substrate is in a rapid equilibrium which is much faster than the catalytic rate. This is an approximation which can be satisfied when the concentration of the ligand is much greater than that of the enzyme. Under these conditions, an approximation is made that the rate of product formation is directly proportional to the fractional occupancy of the enzyme when a zeroth order rate constant for the catalytic step is imposed. When this holds true, the same equation can be used for protein-protein interactions because the same kinetic assumptions apply. Concentration of the titrated protein [B] is substituted for substrate, V is substituted with $\theta = [AB]/[A]$ (fractional occupancy), with $[B] >> [A]$, and $V_{max} = 1$ (complete fractional occupancy). Since we used [A] = 25 nM of FITC-UAP56$_{1\text{-}19}$, which is much less than the apparent K_D of around 600 nM we observed, this experimental setup would produce a valid measurement if all assumptions about binding kinetics were true.

What we see however, is that our curve has far too sharp of an inflexion to be able to mathematically fit this model with any acceptable fitting error. The data do however, fit the Hill equation nicely. The Hill equation has the same underlying assumption, since the Michealis-Menten equation is in fact a special case of the Hill equation, except that the latter does not assume that only one binding event can occur simultaneously for each species and allows for cooperativity or anti-cooperativity of binding[11]. In this case, the fractional occupancy can be written as $\theta = 1 /((K_{eq} / [B])^n + 1)$. Our data fit the Hill equation with a K_{eq}, the concentration of half-occupancy, being around 600 nM and a Hill coefficient (n) of about 2.5. This indicates an

apparent cooperativity of binding as manifested in the steepness of the inflection in our binding curve.

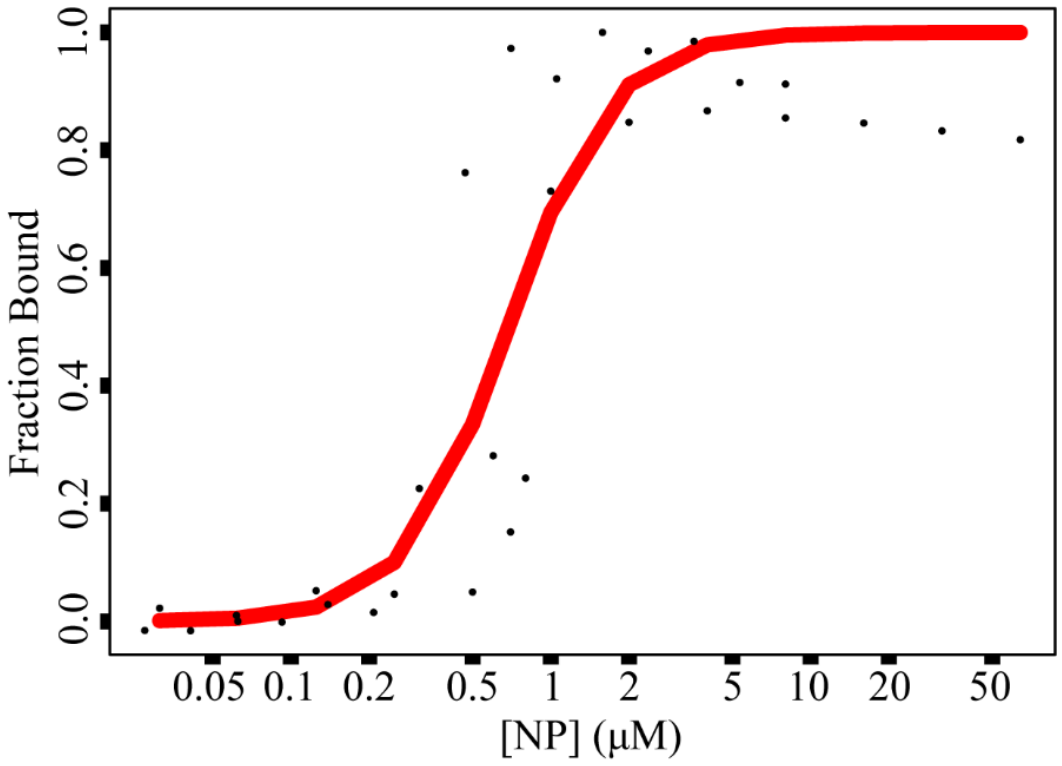

Figure 2.12 – Microscale Thermophoresis of FITC-UAP56-NTE$_{1-19}$ with NP*

Thermophoresis measured for 25 nM of FITC-UAP56-NTE$_{1-19}$ with serial dilution of NP*. Fraction bound calculated as the ΔF for each sample normalized to the maximum ΔF. Points shown across two independent replicates with a line of best fit to the Hill equation. K_{eq} was fit as 622 nM with an 86 nM fitting error. There were insufficient replicates to quantify experimental error.

2.3.3.6 – Fluorescence Anisotropy Corroborates MST

A similar experiment was also carried out using fluorescence anisotropy (FA). In part because MST data had a higher signal to noise than desired, and also to examine if the apparent

cooperativity seen was an artefact of MST, perhaps due to adsorption to the capillary or denaturation caused by heating. FA operates on the same physical principle as MST but measures rotational diffusion rather than translational diffusion. The fluorescent reporter species is hit with plane polarized light at the excitation frequency of the fluorophore. Small molecules undergo enough tumbling in solution over the lifetime of the fluorophore excited state that their emission is in an effectively random orientation, and so the light detected in the plane parallel to excitation and perpendicular to excitation are nearly equal. Large molecules tumble slowly enough that the polarization of their emission is anisotropic, with the parallel component dominating. As with MST, the formation of a complex leads to a change in effective mass and size, and so changes the rotational diffusion or correlation time of the fluorescent reporter. So, when a fixed amount of fluorescently labeled protein is titrated with its binding partner, the same type of binding curve is generated[199].

When the same manner of titration with a low fixed concentration of FITC-UAP56-NTE$_{1\text{-}19}$ and a varying amount of NP* was performed, the curve for the anisotropy data looked very similar to that of MST (**Fig. 2.13**). In the same manner as for the MST data, the points could not be fit to the Michealis-Menten equation but fit well to the Hill equation. The curve had considerably better signal to noise, but very close parameters with a K_{eq} of about 650 nM and a Hill coefficient of close to 2.

Together, MST and FA give orthogonal validation to the less biophysical methods of native EMSA and co-precipitation. They also agree well with each other in their fitting parameters and offer some quantification to the UAP56-NTE NP interaction. However, any quantitation from the fluorescence techniques is rough, as insufficient replicates (2 for MST and 1 for FA) were taken

to get an idea of the experimental error. The fact that the same cooperativity of binding is seen by FA as seen for MST lends more credence to the cooperative binding model, however both methods operate on the same set of physical principles and could thus both be subject to the same types of artefacts. Currently we still lack an adequate explanation for the mathematical cooperativity of binding which the data fit to. It is possible, perhaps, that under certain circumstances or at higher concentrations some multimerization of NP* and WSN NP R416A might be occurring which could explain the cooperativity seen. Perhaps binding to the NTE facilitates this, or maybe multiple NP* can bind to a single UAP56-NTE? If multimerization was occurring it would be at odds with the SEC profile of NP*, which even at the high concentrations loaded onto the column, migrated uniformly at a monomeric size (see **Fig. 2.4 D**). We have insufficient data to claim that this is occurring, or to say that the apparent cooperativity seen is or is not an artefact and cannot offer a physical model that explains this pattern of binding currently.

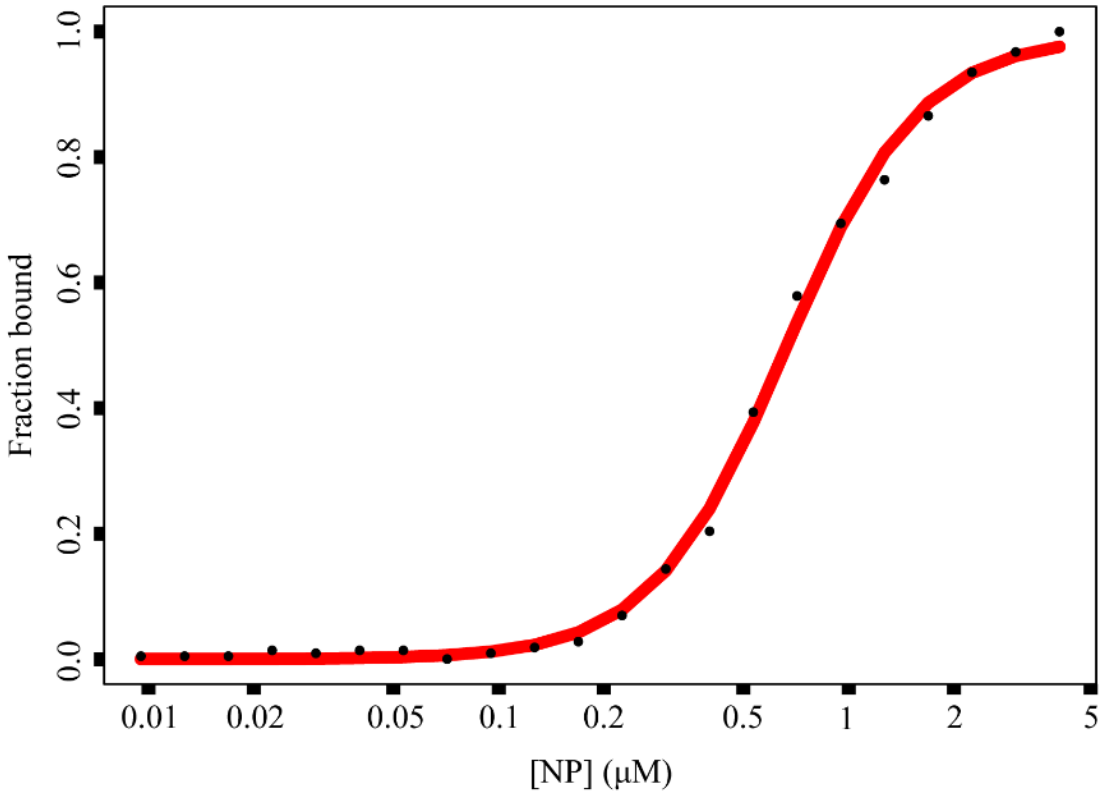

Figure 2.13 – FA of FITC-UAP56-NTE$_{1-19}$ with NP*

Anisotropy was measured with 40 nM of FITC-UAP56-NTE$_{1-19}$ with serial dilution of NP*. Fraction bound was calculated as the difference in anisotropy between each sample and 40 nM FITC-UAP56-NTE$_{1-19}$ without NP* and normalized to the maximum value. Points from a single replicate were fit to the Hill equation to produce a curve. K_{eq} was fit as 669 nM with an 11 nM fitting error.

2.3.3.7 – Native EMSAs with GFP-UAP56 Compare Affinities of NTE vs FL

For further comparison between the relative strength of binding of UAP56-NTE vs UAP56-FL towards NP*, we utilized the fusion constructs of GFP to the NTE (GFP-UAP56-NTE$_{1-19}$) and the full length (GFP-UAP56-FL). Native EMSAs were performed as had been done previously but visualized using a fluorescence scanner. The GFP tag is critical for a titration since UAP56 bands are difficult to visualize by either Coomassie or silver stain below 1 μM, and we had the

data to indicate that our K_D or K_{eq} was well below level. The fluorescence of GFP allowed for visualization of the UAP56 band down to the low concentration suitable for titration of sub-micromolar binding. For GFP-UAP56-NTE_{1-19}, there are two discrete bands, one lower for the free NTE, and the other higher band for the NTE-NP* complex (**Fig. 2.14 A**). In a similar manner to the behavior of the binding curve seen in MST and FA, the shift between these bands occurs rather abruptly, although there is a smear between them indicative of the complex dissociating as it is running. GFP-UAP56-FL did not run as cleanly (**Fig. 2.14 B**). The unbound UAP56 forms a discrete band, but no discrete band is seen for the complex. Instead, the UAP56 band spreads increasingly upwards into a diffuse smear. The reason for the difference in migration of the GFP-UAP56-NTE_{1-19} NP complex and the GFP-UAP56-FL is not readily apparent. However, the shift between free and bound states for both seems to occur between 300 and 400 nM. These native EMSAs are in generally good agreement with the coprecipitation experiments in demonstrating that the NTE is by far the biggest contributor to NP binding and that in isolation, UAP56-NTE is capable of binding NP at near WT levels.

The GFP fusion constructs were also used for MST to try to quantitatively assess differences in binding between UAP56-NTE and UAP56-FL. Despite several attempts, this approach was hampered by several problems. The limited runs which produced usable data were in qualitative agreement with previous data. A binding curve could be produced, and for both GFP-UAP56-NTE_{1-19} and GFP-UAP56-FL this interaction was sub-micromolar. No dramatic difference in affinity was seen between the constructs. We put limited credence to any conclusion that might be drawn from this though, as signal-to-noise ratios were low, unconventional properties of the time traces required differing analyses, and most importantly reproducibility was difficult.

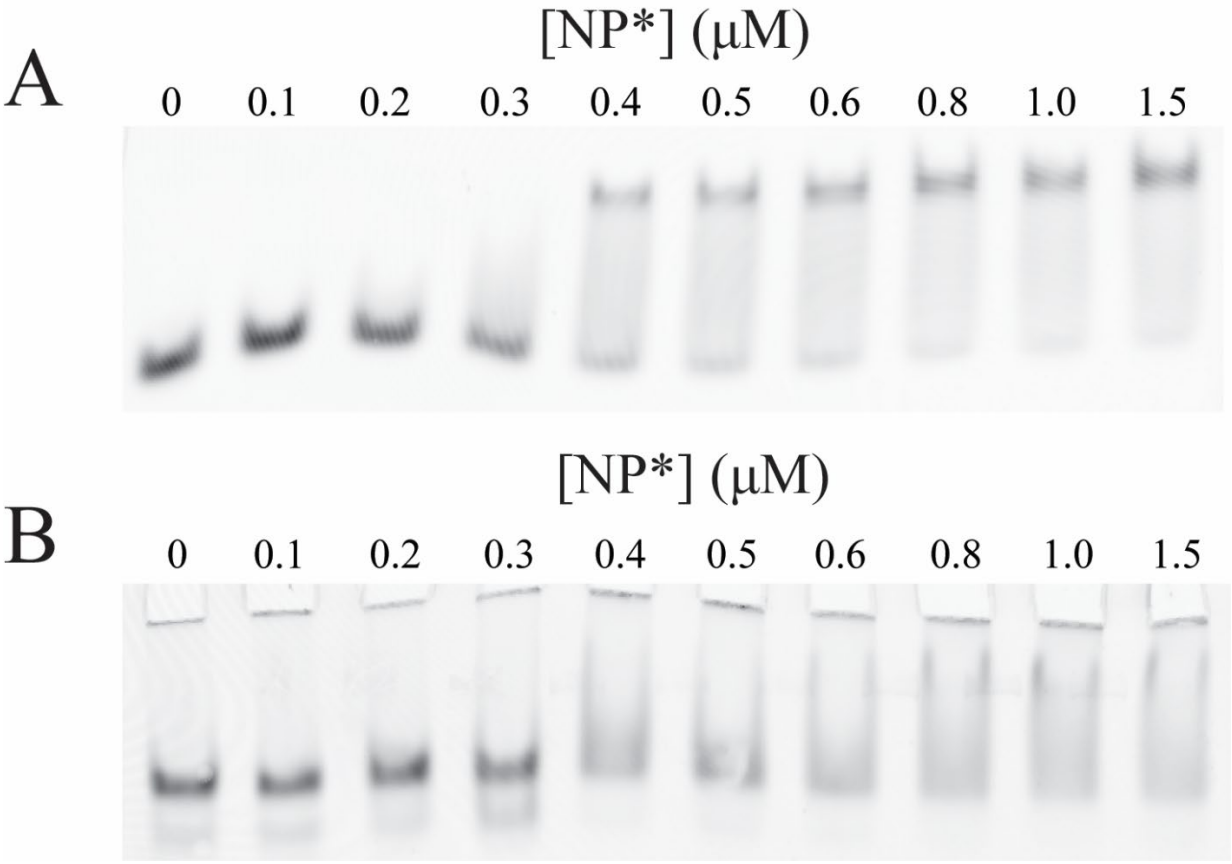

Figure 2.14 – EMSA Titration of GFP UAP56 Estimates Relative Affinities

A) GFP-UAP56-NTE$_{1-19}$ at a constant 100 nM was titrated with increasing amounts of NP* and sample run on native PAGE and visualized by fluorescence. The lower band is the free UAP56, while the higher band is the complex. **B)** Same as A, except using GFP-UAP56-FL. In this instance, the lower band fades, but the complex band does not form cleanly. The faint band under the lower free UAP56 band is a degradation product.

2.4 – Discussion

The initial part of this study focused on first reproducing what was previously seen by others in the literature. It had been previously established that UAP56 and NP interact by a yeast two hybrid screen and coprecipitation[187]. We could reproduce this finding not only with NP from the previously used PR8 strain, but also with NP from the WSN strain of IAV. The initial paper to

identify this interaction, claimed the minimal binding region for UAP56 lies in the C-terminal RecA domain. We were able to see some interaction between UAP56-RecA-C and NP at high concentrations, but this was by native EMSA which is permissive to non-specific binding. By coprecipitation, binding was seen with these RecA domains together as GST-UAP56-Core, which agrees with the first study, however this binding was weak compared to GST-UAP56-FL. Importantly, we identified a region of UAP56 not previously reported to be involved in this interaction, and showed that this NTE which we identified is the principle element responsible for NP binding as it can co-precipitate with similar efficiency to wild type. Of note, the NTE region was excluded from the coprecipitation experiment in the original paper. We further narrowed down the region in the NTE binding element to be a conserved patch of acidic residues in amino acids 1-19. A more recent paper claimed that UAP56 interacts only with NP in a trimeric form[190]. We showed that a monomeric point mutant of NP, R416A, can indeed interact with UAP56 and can also interact with the NTE in isolation. This interaction, as well as that of the WSN strain, is more sensitive to disruption by salt, which may explain why it was missed previously. Fusions of the NTE and full length UAP56 to GFP showed that they have similar binding when examined by native EMSA, and MST and FA show in a rough quantitative manner that this interaction is in the hundreds of nanomolar range for NP*, which further confirms the ability of the monomer to bind to UAP56. Curiously, the fluorescence methods also demonstrated that there is an apparent and unexplained cooperativity of binding for NP* towards the UAP56-NTE. We do not have a model for this, but it may represent an interesting aspect for others to investigate in future biophysical studies.

CHAPTER 3

THE UAP56-NTE RECOGNIZES THE NUCLEIC ACID BINDING SITE OF NP IN COMPETITION WITH RNA

3.1 – Introduction

The first part of this study focused on finding the minimal components needed for NP binding to UAP56, and how binding was influenced by NP oligomerization and buffer conditions. In the second half of this study, which is described in this chapter, we built on the experimental findings of the first chapter to explore this interaction in detail at the atomic level. Obtaining a crystal structure of the UAP56 NP complex would be the most informative way of revealing how the interface between them is formed, and so made up our initial structural approach. The elucidation of the location and critical residues for the interaction were, in the end, obtained by crosslink mass spectrometry and tested through mutagenesis. Here is described how these methods were used to extract more detailed information about the interaction initially described in chapter one, and from this, the influence and functional implications of RNA on binding.

3.2 – Methods

3.2.1 – Crystallization of NP

WSN NP was crystallized with 70 ul reservoir solution containing 10% glycerol, 10% PEG 8000, 0.2 M Na Acetate pH 6.6. 250 nl of 55 μM protein in 10 mM HEPES, pH 7.0, 200 mM

NaCl, 0.5 mM TCEP was mixed with 250 nl of reservoir solution. WSN NP R416A was crystallized with 70 ul reservoir solution containing 100 mM HEPES pH 7.5, 1.2 M Na K tartrate, 3% isopropanol. 200 nl of 100 μM protein in 10 mM HEPES, pH 7.0, 50 mM NaCl was mixed with 200 nl of reservoir solution. WSN NP Δ402-429 was crystallized with 70 ul reservoir solution containing 100 mM Tris HCl pH 8.5, 18% PEG 3350, 10% glycerol, 10 mM DTT, 2% w/v benzamidine HCL. 200 nl of 90 μM protein in 10 mM HEPES, pH 7.0, 150 mM NaCl was mixed with 200 nl of reservoir solution. All proteins were crystallized by sitting drop vapor diffusion at room temperature in 96 well MRC2 plates set up by TTP mosquito system.

For each NP construct crystallized, a set of un-soaked crystals was harvested, as well as a set harvested after the addition of super-saturating amounts of lyophilized powder of FITC-UAP56-NTE_{1-19} to the drop for 12 hours. For WSN NP R416A and WSN NP Δ402-429, an additional set of crystals was harvested which were soaked for 12 hours with the addition of tag-less UAP56-NTE_{1-38}.

Frozen crystals were shipped to the Advanced Photon source and diffracted remotely on the LS-CAT, 21-ID-D beamline. Diffraction data was processed in HKL2000[200]. Phasing for each dataset was done by molecular replacement using as search models PDB: 3ZDP for WSN NP R416A[154], PDB: 4IRY for WSN NP Δ402-429[192], and PDB: 2IQH for WSN NP WT[153]. Phasing and model refinement was performed in Phenix[201].

3.2.2 – FITC-UAP56-NTE1-19 Crosslinking to NP*

FITC-UAP56-NTE_{1-19} and NP* were mixed at 0.5 μM and 1 μM respectively in the reaction buffer (10 mM HEPES pH 7.0, 150 mM NaCl, 0.5 mM TCEP, and 5 mM 1-Ethyl-3-(3-

dimethylaminopropyl) carbodiimide Hydrochloride (EDC)), and incubated for 1 hour at room temperature. The reaction was quenched by the addition of Tris pH 8.0 to 100 mM. Efficient crosslinking was verified by SDS-PAGE and fluorescence imaging by Typhoon FLA 9000 biomolecular imager to observe the size shift of the UAP56 band.

3.2.3 – Trypsin Digestion and Liquid Chromatography Tandem Mass Spectrometry (LC-MS/MS) †

After EDC crosslinking, a 10 µL aliquot of the sample was mixed with 10 µL of 10 M urea. TCEP was added to a final concentration of 2.5 mM. The sample was incubated at 25 °C for 30 min followed by addition of iodoacetamide to a final concentration of 5 mM and was incubated at 25 °C in the dark for 45 min. The sample was then diluted with 50 mM Tris (pH 8.0) to 100 µL. 0.1 µg of trypsin was added to the sample and incubated for 18 hours at 37 °C. 0.4 µg of endoproteinase Asp-N was then added to the sample for overnight digestion at 37 °C. The sample was dried and reconstituted in 10 µL of 0.1% formic acid and desalted using a ZipTip C18 pipette tip. The sample was reconstituted in 10 µL of 0.1% formic acid and 5 µL was used for LC-MS/MS analysis.

The 5 µL sample was separated over a fused silica capillary HPLC column (250 mm x 100 µm) packed with Phenomenex Jupiter resin (3 µm mean particle size, 300 Å pore size) on an UltiMate 3000 RSLCnano System. A 70-minute gradient was performed, consisting of the following: 0-60 min, 2-45% ACN (0.1% formic acid); 60-70 min, 45-95% ACN (0.1% formic acid) balanced with

† Protein fragmentation and all mass spectrometry steps, and the interpretation of the data were performed by Zhen Wang, Ph.D. in the lab of Kevin Schey, Ph.D., who also wrote this method which was taken from Morris *et al*, 2020[3], on which they are co-authors.

0.1% formic acid. The eluate was directly infused into a Q Exactive instrument equipped with a nanoelectrospray ionization source. The data-dependent instrument method consisted of MS1 acquisition (R=70,000), using an MS AGC target value of 1e6, followed by up to 15 MS/MS scans (R=17,500) of the most abundant ions detected in the preceding MS scan. The MS2 AGC target value was set to 1e5, with a maximum ion time of 100 ms, and intensity threshold of 3e4. HCD collision energy was set to 28 and dynamic exclusion was set to 10 s.

To identify residues that are involved in crosslinking between NP* and FITC-UAP56-NTE$_{1-19}$, the raw data were converted to MGF file by Scansifter[202], a tool under development at Vanderbilt University Medical Center and crosslinked peptides were then searched by StavroX, version 3.6.6[203]. All output candidate peptides that have scores above 60 were manually checked to verify crosslinked peptide assignments. Crosslinked peptides were confirmed when a series of b-or y-ions were present from both peptides and they matched the most intense fragment ion signals. FITC-UAP56-NTE$_{1-19}$ does not have trypsin cleavage sites. Therefore, it can be cleaved only by Asp-N. Expected peptides from FITC-UAP56-NTE$_{1-19}$ after *in silico* digestion by Asp-N with a maximum of two missed cleavage sites were obtained using Protein Prospector. Met 1 oxidation was considered as variable modification. This list of predicted Asp-N peptides and the mass shift, if each of these peptides was involved in crosslinking, can be found in the Table S1. Crosslinked peptides were also searched using TagRecon[204] by searching modifications on the NP protein with each of UAP56 peptides generated by Asp-N. MS/MS fragments were then used to manually confirm the crosslinked peptide as described above.

3.2.2 – Native EMSA Titrations Examining FITC-UAP56-NTE$_{1-19}$ or RNA Binding to NP* Mutants

All binding titrations were conducted in a buffer containing 10 mM Tris pH 8.0, 150 mM NaCl, 0.5 mM TCEP, 5% glycerol. For titrations of NP* or mutant to synthesized N-terminal FITC labeled UAP56-NTE$_{1-19}$, 0.1 μM UAP56 peptide was incubated with varying amounts of NP* or mutant for 20 min at room temperature. For RNA binding titrations, 15 nM of synthesized 15-mer poly(U) ssRNA with Alexa488 at the 5' end was incubated with varying amounts of NP* or mutant in the presence of 0.5 U/ul SUPERase-In for 10 minutes at room temperature.

For the RNA-UAP56-NTE competition assay, a mixture of 0.1 μM FITC-UAP56-NTE$_{1-19}$ and 1.5 μM NP* was incubated with varying amounts of synthesized 15-mer poly(U) ssRNA in the presence of 0.5 U/ul SUPERase-In for 10 minutes at room temperature. In all EMSAs, 10 ul of each sample were loaded on a 5% native polyacrylamide gel that was prepared with 45 mM Tris and 45 mM boric acid and pre-run in the same buffer. Gels were run for 30 minutes at 100V at 4 °C. All gels were visualized by Typhoon FLA 9000 biomolecular imager. Densitometry was performed using ImageJ[205]. For each lane, the integrated intensity of a 6.6 mm x 3.3 mm box centered on the upper band corresponding to the complex was taken. For each gel, the integrated intensities were normalized to the most intense band on the gel. Normalized intensities were then averaged across the three replicates.

3.3 – Results

3.3.1 – Crystallization of Several NP Variants could not Resolve Binding Interface

NP had been previously well characterized structurally. The first crystal structure was of WSN strain NP, which was crystallized as a trimer[153]. Since then, many variants on this structure have been determined, including in complex with antibodies and small molecules[154,192,206–208]. Our other molecule of interest, UAP56, also has an available crystal structure[66]. Given that we knew the two formed a complex and that individually both species had been crystallized, the possibility was opened that they could be co-crystallized. Co-crystallization is, however, very technically challenging. Even for highly purified and well-ordered proteins that crystallize easily on their own, the protein complex that they form may not pack well together in a crystallizable manner. If it is possible to co-crystallize a complex, it may require a different set of conditions than either of the components on their own and might require beginning a tedious screening process over again from the ground up. While such a task was not completely out of the question from the onset of the project, to attempt to do so would have been overly ambitious, most likely. Compounded to this was the issue of purity for UAP56. As previously mentioned, degradation was an issue in UAP56 purification, and proteolytic fragments of our various constructs proved very difficult to eliminate completely. While minor contaminants are acceptable for in vitro studies, they would have likely made crystallization very challenging.

When we identified UAP56-NTE$_{1\text{-}19}$ as the minimal binding region for NP, we saw an opportunity to use a small fragment of UAP56 for crystallization rather than UAP56-FL. Crystallization of protein fragments or individual domains of proteins is generally considerably more likely to succeed in producing atomic resolution structures than larger multidomain

constructs because of the reduced conformational states that can be occupied and elimination of disordered regions. Another advantage is that UAP56-NTE$_{1-19}$ is a peptide, and thus can be synthesized. This eliminates problems of purity for UAP56 which we had experienced. Additionally, as a peptide, its properties are between a that of a small molecule ligand and that of a folded protein. Co-crystallization of a protein, especially one such as NP with an existing apo structure, with a small molecule ligand is usually much more tractable than protein-protein co-crystallization. Small molecules are often capable of being soaked into a crystal, owing to their diffusibility and ability to move through solvent channels. At 2 kDa, it is quite large for a small molecule, but even larger peptides have been soaked into crystals successfully before. For these reasons, we pursued a strategy of crystallizing NP in the apo form and soaking in FITC-UAP56-NTE$_{1-19}$ to obtain an atomic structure of the two in complex, as we believed this to be the most tenable approach.

We began by crystallizing WSN NP WT. We tested the crystallization conditions based around those described in the original paper describing the structure [153]. Grid screens of a gradient of concentrations of PEG 8000 versus varying pH using sitting drop vapor diffusion produced the optimum conditions for crystal formation. We further tested several additives in screens to control nucleation and produce singular crystals as large as possible. Our final conditions were very close to those described in the original publication. We then harvested several crystals without soaking and collected a set of crystals in which FITC-UAP56-NTE$_{1-19}$ was soaked into the crystals over 12 hours by addition to the mother liquor. The FITC label at the N-terminus of the peptide allowed for monitoring of soaking, as it colors the mother liquor yellow, and when sufficiently soaked into the crystal, the crystal would change color (**Fig 3.1 A**). The harvested crystals for both soaked and un-soaked datasets were then diffracted at a synchrotron beamline.

Phasing for both datasets was done by molecular replacement using the published WSN NP as a search model. In both cases the resulting model after automated refinement was almost indistinguishable from the published WSN NP structure. A difference density map showed no major region that would be consistent with FITC-UAP56-NTE$_{1-19}$. So, we could reproduce the published structure but could not see the NTE anywhere.

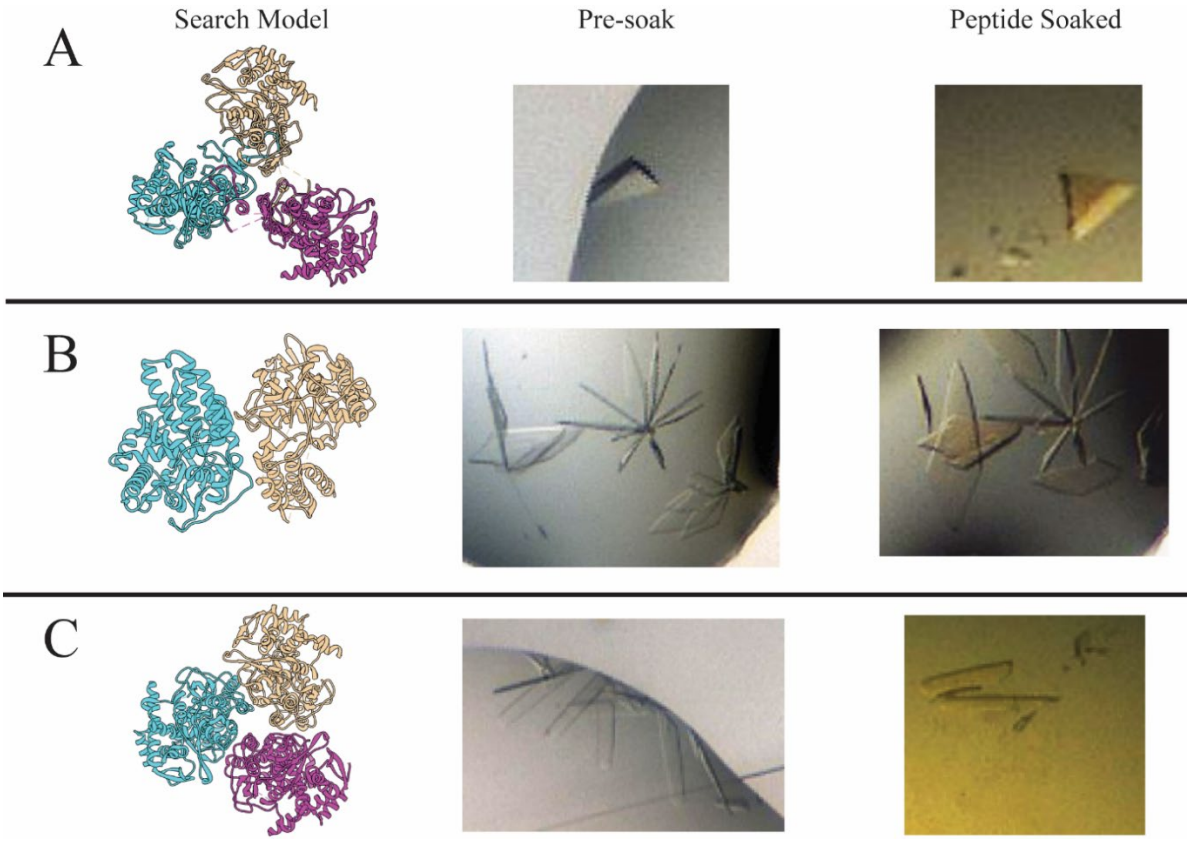

Figure 3.1 – X-ray crystallography of NP with UAP56 NTE peptide

Left panels show the published crystal structures used for phasing for **A)** WSN NP WT (PDB: 2IQH)[153], **B)** WSN NP-Δ401-429 (PDB: 4IRY)[192], and **C)** WSN NP R416A(PDB: 3ZDP)[154]. The right panels shows the same crystallization drop as in the center panels after a 12-hour soak with FITC-UAP56-NTE$_{1-19}$ before harvesting, which were used for diffraction. Note that WSN NP-Δ401-429 crystallized with different geometry from the other two and in a different space group. Models displayed with UCSF Chimera[69].

From here we reasoned that resolution may be the problem. We could see that the crystal had soaked successfully, but perhaps we had insufficient resolution to discern it. Our dataset for WSN NP had 4 Å resolution in its highest shell, which is insufficient to make out most side chains. Even if we harvested and diffracted again to try to obtain higher resolution, it is unlikely we could get higher resolution than the published structure which itself which was at 3.3 Å[153]. Even at that resolution, making residue assignments for a peptide would not be trivial, and it would be difficult to see which residues between NP and UAP56 were interacting, which was the ultimate goal. There are, however, crystal structures of mutants of WSN NP which have higher resolution. A mutant in which the flexible tail loop by which NP multimerizes has been deleted (WSN Δ402-429) was crystallized and diffracted to 2.8 Å[192]. Another mutant, WSN R416A, the same residue as in NP*, diffracted to 2.7 Å[154]. To obtain higher resolution we grew, expressed and crystallized both of these mutants.

WSN NP R416A and WSN Δ402-429 purified in essentially the same way as NP*, but fractions were cut more stringently for higher purity. As with WSN NP WT, WSN R416A crystallization began as a grid screen around the published condition varying combinations of pH and precipitant to that which was most conducive to crystallization in our hands. As before, an additive screen identified that 3% isopropanol helped to control nucleation to produce larger singular crystals for diffraction. Unlike WSN R416A, WSN Δ402-429 did not crystallize at all under when screened around the conditions reported in the initial publications. However, the addition to the reported crystallization conditions of 2% benzamidine HCl, which we identified through a screen, led to large plate crystals which formed in clusters.

To maximize our chances of success, we performed the soak using the FITC-UAP56-NTE$_{1-19,}$ which was chemically synthesized, as well as UAP56-NTE$_{1-38}$ which was purified as a GST fusion and cleaved from the tag, with GST and GST-TEV recaptured over glutathione resin. The exact purity and concentration of UAP56-NTE$_{1-38}$ was difficult to ascertain given that the extinction coefficient for a peptide with no tryptophans is small and an unreliable prediction. Additionally, because it is highly acidic and small, it neither produces significant signal by the Bradford colorimetric assay, nor does it stain well on a gel, although it could be visualized by silver stain when enough was loaded. The larger size of UAP56-NTE$_{1-38}$ compared to FITC-UAP56-NTE$_{1-19}$ meant that soaking was less likely to be possible, however without fluorescence tag, we could not know if soaking worked until diffraction. Despite not having perfect purity, larger protein contaminants would not likely cause an issue as they should not be able to soak in, and while the concentration had considerable uncertainty in measurement, the peptide was added well above what should be saturating conditions for soaking.

We soaked FITC-UAP56-NTE$_{1-19}$, harvested, and diffracted these two mutants in the same manner done for WSN WT. Both were also phased in the same manner using their respective published crystal structures as search models (**Fig. 3.1 B** and **C**). Unfortunately, in both these cases the result was the same as for WSN WT in that neither produced any difference density that could be attributed to the bound NTE. In addition, the WSN Δ402-429 dataset diffracted to 2.8 Å, a resolution which should have been able to discern this density if it were there. From this we reasoned that crystallography would not be able to give us molecular detail about the specifics of the UAP56 NP interaction, and so we decided to pursue answers by a different methodology.

3.3.2 – Crosslink Mass Spectrometry Reveals UAP56-NTE and NP* Residues in Close Contact

Without crystallography to give us atomic detail, we turned to another method that can provide some of the same information. Crosslink mass spectrometry (XL-MS) is one method to identify which residues across multi-protein complexes form an interface. There are a variety of choices of chemical crosslinkers available which let one select what type of chemical moiety will link to what, and the proximity the residues must be in to form the link. Crosslinkers with broad specificity such as glutaraldehyde are common when the interacting partners are not known and when one wishes to look at all protein-protein interactions. It is notorious however, for producing many false positives. To minimize the detection of these, we chose a crosslinker that only reacts to two specific moieties. The chemical EDC will only react with primary amines on one end and carboxylic acids on the other (**Fig. 3.2 A**). This means that it will only link lysines or the N-terminus to aspartates, glutamates, or the C-terminus. The crosslink process leaves behind an amide which links the two, and no intervening spacer. Because EDC is a zero-length crosslinker, amines and carboxylic acids will only react if they are within several Å of each other. PR8 NP has 21 lysines, with most clustered in and around the RNA binding cleft while UAP56-NTE$_{1\text{-}19}$ has 10 total aspartates plus glutamates (**Fig. 3.2 B**). This makes EDC an ideal crosslinker, as these moieties are quite abundant. Because we suspect the UAP56 NP interaction is likely charge-charge driven, some of them are likely in close contact forming salt bridges.

When the crosslinking was performed, it was quite efficient. The product was run on SDS-PAGE and the fluorescence of FITC was used to identify the crosslinked FITC-UAP56-NTE$_{1\text{-}19}$, which ran at a high molecular weight (**Fig. 3.2 C**). Compared to unlinked control, which ran at the bottom of the gel, a significant portion of the protein formed intermolecular crosslinks.

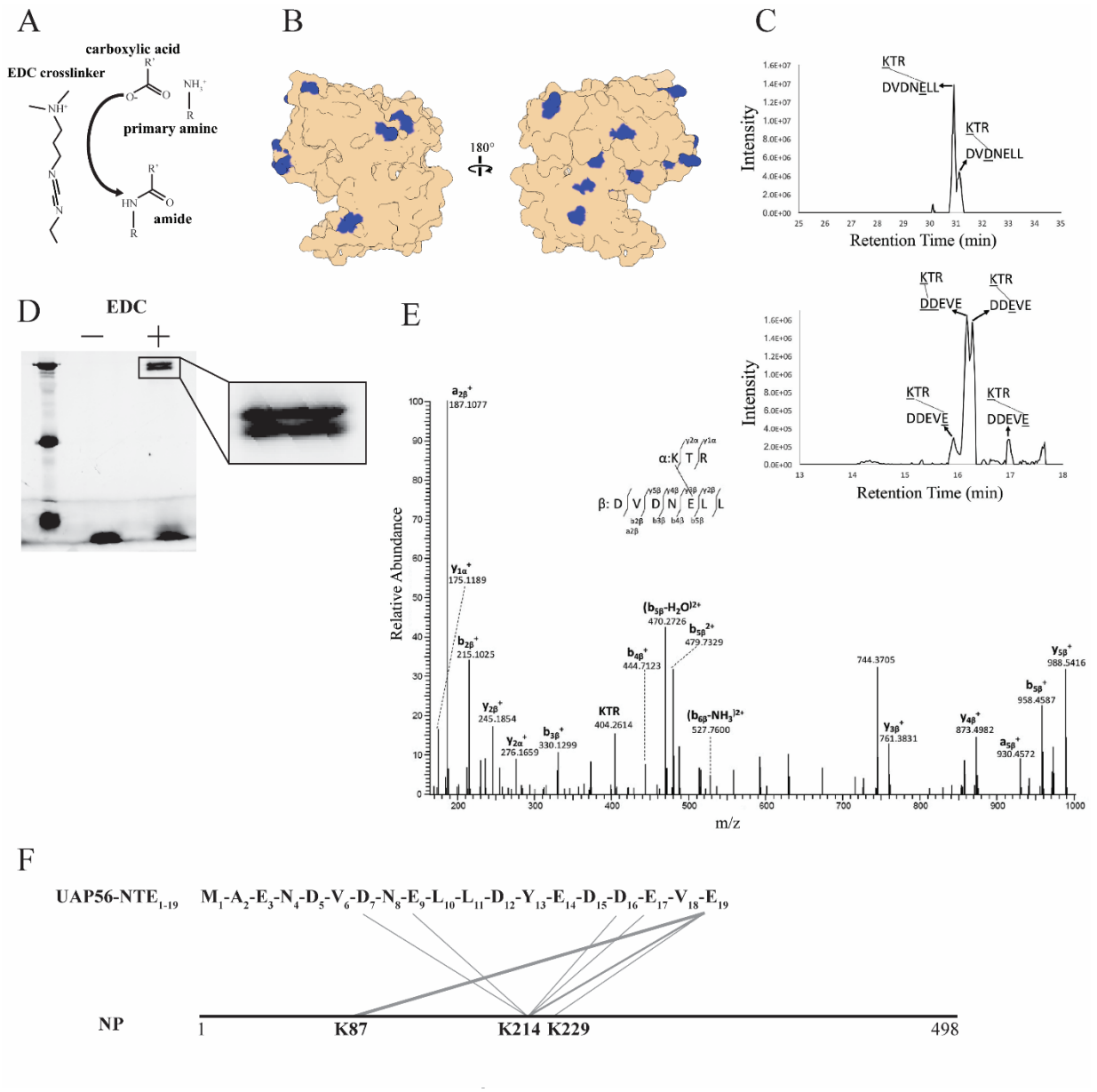

Figure 3.2 – XL-MS Reveals UAP56-NTE and NP* Residues in Close Contact

A) Schematic showing the chemical structure and mechanism of the EDC crosslinker utilized. **B)** Surface representation of WSN NP (PDB: 3ZDP) showing the location of solvent accessible lysine position on PR8 NP in blue which could be crosslinked by EDC[153]. **C)** Selected ion chromatograms for crosslinked peptides involving K214 of NP* **D)** Fluorescence scan of an SDS-PAGE gel of FITC-UAP56-NTE$_{1-19}$ and NP* in the presence or absence of EDC crosslinker. Note that the crosslinked product runs as a doublet. **E)** Tandem mass spectrum for crosslinked peptide K*TRDVDNE*LL. **F)** Summary of crosslinks identified by the mass spectrometry approach along the sequence of the NTE and NP*. The width of the line corresponds to the number of hits for that residue pair. Panels C), E) and F) were taken from Morris *et al*, 2020[3].

Curiously, the band of linked FITC-UAP56-NTE$_{1-19}$ formed a doublet. The two bands very close in size might represent two versus one FITC-UAP56-NTE$_{1-19}$ crosslinked to NP*, but not differing numbers of NP*. We tried varying the ratio of NP to NTE from 4:1, 1:1, and 1:4, but could not eliminate this doublet, and so ultimately, we moved forward with a 1:1 ratio.

‡ Digestion of the crosslinked complex with trypsin and Asp-N endopeptidase created peptide fragments. Given the possible sites Asp-N and trypsin can cut, all possible combination of fragments, and their corresponding masses can be calculated. When fragments were separated over HPLC and subject to electrospray ionization and mass spectrometry as they eluted, we were able to determine the fragments that were crosslinked. When tandem mass spec was incorporated, this allowed the exact residues on NP* to be mapped to the exact residues on UAP56-NTE$_{1-19}$ to which they were crosslinked. Two residues on NP* were shown to crosslink with UAP56-NTE$_{1-19}$, K214 and K87. Crosslinks were made from these points to D7, E9, D16, E17, and E19. E19 was by far the most frequent crosslink site, and K87 was only detected as crosslinking to this residue. This may be because it is also the C-terminus, which is itself cross linkable, or because as the terminal residue it is likely the most flexible. One crosslink from E19 of the NTE to was detected to K229 of NP*, however, it was not picked up when the computer analysis was repeated a second time and is likely insignificant.

‡ Protein fragmentation and all mass spectrometry steps, and the interpretation of the data were performed by Zhen Wang, Ph.D. in the lab of Kevin Schey, Ph.D., both of whom are credited as authors on the publication produced[3].

3.3.3 – Mutation of Crosslink-Identified NP Residues has Minimal Impact on NTE Binding

The mass spectrometry identification puts UAP56-NTE$_{1-19}$ in the vicinity of the RNA binding cleft. The two residues identified by XL-MS, K87 and K214, sit on the periphery of this groove. K214 is a highly exposed residue at the edge of the cleft in the head domain. K87 lies in a flexible loop which projects outward from the bottom rim of the cleft on the body domain opposite K214 (**Fig. 3.3 A**). Since these residues make direct contact with the NTE, and would likely create a salt bridge, we reasoned that they might be critical for binding. We created a charge swap mutant K214D (NP*-K214D). We also made a mutant targeting K87, in which the loop it resides on, residues 74-88, are replaced with a glycine-serine linker (NP*-ΔLoop). Once expressed and purified, both NP*-ΔLoop and NP*-K214D were examined for their ability to bind FITC-UAP56-NTE$_{1-19}$. This was done through a native EMSA similar to those performed using the GFP constructs. A low and fixed amount of FITC-UAP56-NTE$_{1-19}$ runs as a distinct band into the gel in each lane, and the formation of a complex with increasing amounts of either NP*, NP*-K214D, or NP*-ΔLoop is seen as the slower running band which barely enters the well (**Fig. 3.3 B**). Compared to NP*, NP*-ΔLoop shows no significant change in migration pattern. Both have an upper complex band forming between 200 and 400 nM, with free FITC-UAP56-NTE$_{1-19}$ essentially totally depleted at 1.6 μM. NP*-K214D did show a minor difference in binding affinity, with the free FITC-UAP56-NTE$_{1-19}$ band not approaching complete depletion until the end of the titration range at 6.4 μM. While weaker than NP*, it was not a dramatic effect. If K214D formed a critical salt bridge, it would be expected that binding would be almost completely abolished.

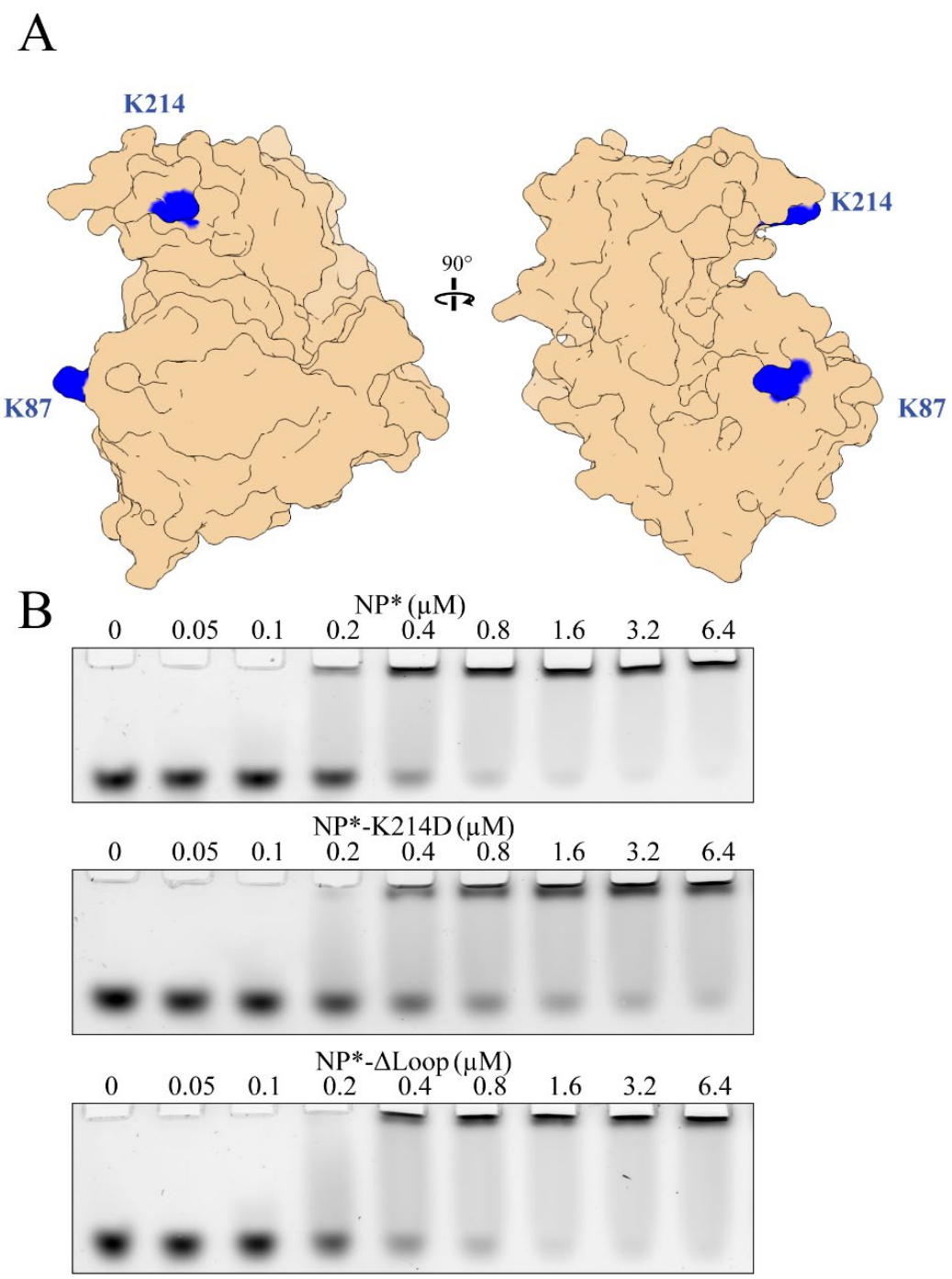

Figure 3.3 – Mutation of Crosslink-Identified NP Residues has Minimal Impact on NTE Binding

A) Highlighted location of identified crosslinked residues superimposed on WSN NP R416A structure (PDB: 3ZDP)[154]. Figure was generated with UCSF Chimera[69]. **B)** Native EMSA of 100 nM FITC-labeled UAP56-NTE$_{1-19}$ binding to increasing amounts of NP*, NP*-K214D, or NP*-Δloop, visualized by fluorescence. The lower bands are free NTE, and bands at the top of the gel are the NTE NP complex. All experiments were repeated three times independently. Figure was adapted from Morris *et al*, 2020[3].

In conjunction with the NP*-ΔLoop results, these data suggest that the lysines which we identified in XL-MS which are found along the periphery of the binding groove play little role in driving the interaction or are perhaps auxiliary. Of note, there are many lysines near the binding groove, so it is possible that their binding roles are redundant. K214 and K87 are also among the most highly solvent exposed residues in the protein, and so it is quite possible that among the lysines in the RNA binding cleft, they were simply the most available to react with the crosslinker. While these results were not what we had predicted, neither they are incompatible with the hypothesis that UAP56-NTE$_{1\text{-}19}$ occupies the RNA binding cleft. These data are also compatible with, albeit not predicted from, the XL-MS which identified them, as that only identified that these sets of residues came in proximity to one another, not that they had any functional significance.

3.3.4 – Mutation of R174 and R175 Severely Abrogates NTE Binding

When we did not see dramatic abrogation of NTE binding with NP*-ΔLoop or NP*-K214D, we were somewhat puzzled. Clearly, based on the fact that both of these sites overhang the RNA binding cleft on opposite ends and that they were the only major sites picked up by crosslinking, then the NTE must be occupying the space between them in the RNA binding cleft. We reasoned that if lysines periphery to the binding cleft were dispensable, then perhaps it is the charged residues in the center of the binding cleft which are the main point of interaction with the NTE.

To test this, we mutated two highly conserved arginines in a short helix running parallel to the center of the path of the binding cleft (**Fig. 3.4 A**). A charge swap to aspartate was made as this

would be expected to reverse the charge-charge interaction we suspected would be occurring here. NP*-R174D/R175D, when expressed and purified was used in the same native EMSA experimental setup as for NP*-ΔLoop, NP*-K214D, and NP*, in which it was titrated against FITC-UAP56-NTE$_{1-19}$ and the formation of complex bands was observed. Compared to NP* the difference in apparent affinity towards the NTE was severely abrogated (**Fig. 3.4 B**). Whereas the bound state for NTE was almost completely populated with addition of 1.6 μM NP*, a significant amount of free NTE was still observed with NP*-R174D/R175D even at the maximum of the titration range at 6.4 μM. When each of the three replicates for the native EMSA of NP*, NP*-K214D, NP*-ΔLoop, and NP*-R174D/R175D were analyzed by densitometry, averaged, and plotted, it was seen that the concentration required for half-depletion of free NTE was around tenfold higher for NP*-R174D/R175D compared to NP* (**Fig. 3.4 C**). While densitometry is a semi-quantitative method at best, it is still obvious the R174D/R175D mutation is deleterious to binding. With its high density of positive charge and location in the very center of the cleft, it is not surprising these arginines are important for NTE binding. It is important, though, to note why these arginines were not detected by XL-MS. Despite their close contact with the NTE, the EDC crosslinker cannot react with the guanidino moiety on the arginine side chain, but only the primary amines of lysines and the N-terminus, despite both being positively charged and often biochemically very similar. Given that this EMSA result suggests that these two arginines are the primary interacting residues on NP, their central location in the cleft explains why the bound NTE would be in close proximity to crosslink with K214 and K87, which straddle opposite sides of the cleft right near the arginines.

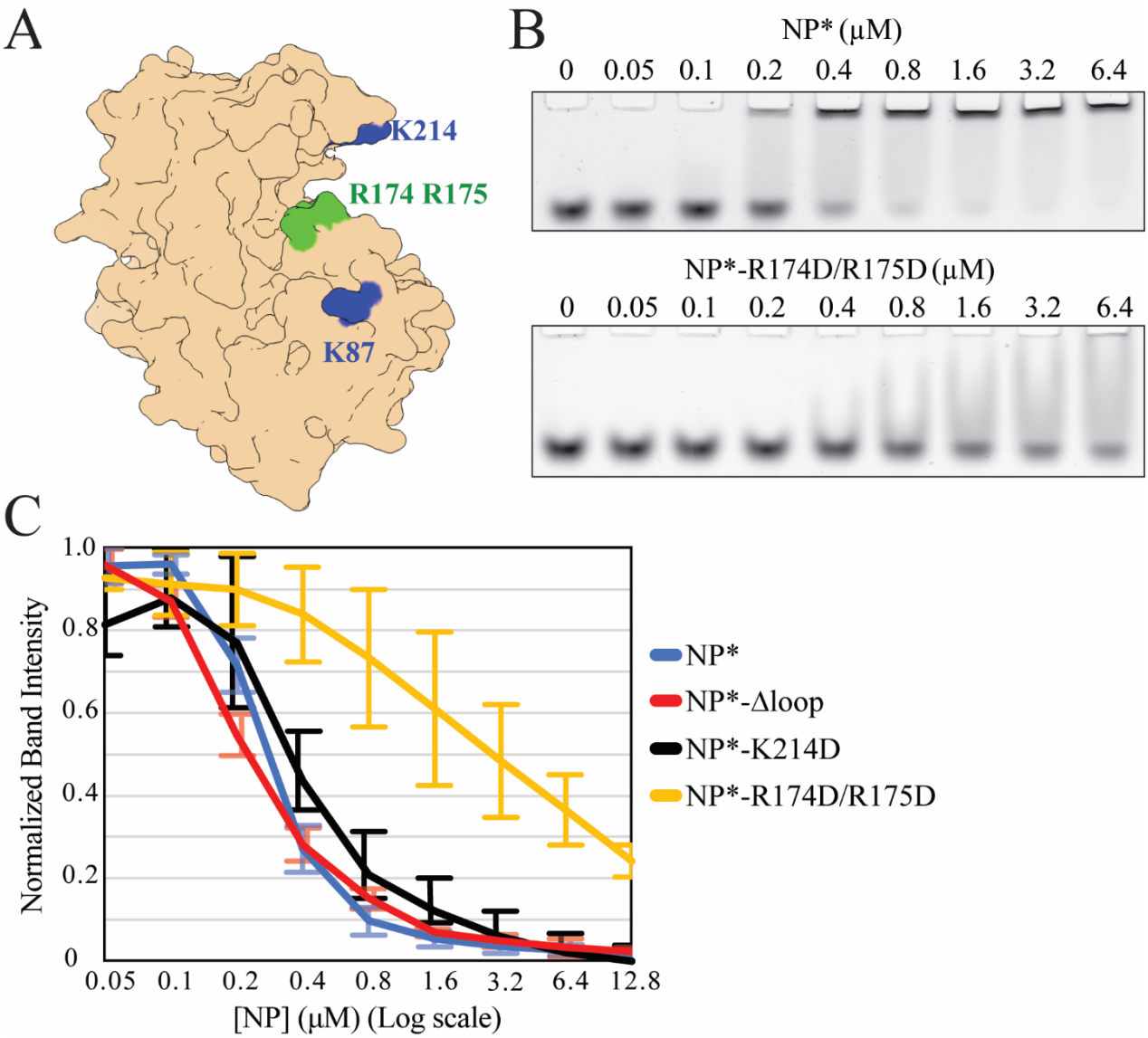

Figure 3.4 – Mutation of R174 and R175 Severely Abrogates NTE Binding

A) Highlighted location of identified crosslinked residues in blue and the R174D R175D mutation made in the center of the cleft in green superimposed on WSN NP R416A structure (PDB: 3ZDP)[154]. Figure was generated with UCSF Chimera[69]. **B)** EMSA assay of 100 nM FITC-labeled UAP56-NTE$_{1\text{-}19}$ binding to increasing amounts of NP* or NP*-R174D/R175D, visualized by fluorescence. For NP*-R174D/R175D, a distinct complex band does not form, but instead runs as a smear. All experiments were repeated three times independently. **C)** Quantification of NTE binding to all mutants tested from the EMSAs presented in B) and Fig. 3.3.3 B). Bars represent standard deviation of band intensity across three images for each mutant. Panel B taken from Morris *et al*, 2020

In this way, the native EMSA results are consistent with the crosslink mass spectrometry data, and together they solidly establish that UAP56-NTE interacts with NP by a charge-charge mechanism at the RNA binding cleft.

3.3.4 – NP* Loop Deletion and R174/R175D Mutations Abrogate RNA Binding

As stated previously, UAP56-NTE$_{1\text{-}19}$ has a high density of negative charge, while NP is highly positively charged. This led us to hypothesize that this was an electrostatically driven interaction. The XL-MS showed us that this was occurring at the charged cleft, and mutagenesis further confirmed this and verified that it was indeed electrostatically driven. RNA also binds to NP in this cleft by an electrostatic mechanism. The obvious implication is that the effects of mutagenesis on the residues we tested should show a similar pattern in abrogation of RNA binding as seen for NTE binding. To test this, we used a very similar setup to that used to test NTE binding. We performed native EMSAs in which the migration of fluorescently labeled RNA was used to assess the affinity of a given mutant towards RNA. With a fixed 15 nM concentration of short fluorescent-labeled RNA, the respective concentrations of the mutant forms of NP* were titrated to create a shift in the RNA band mobility as seen by fluorescence imaging for the NP-RNA complex. This was compared to the migration pattern for NP* (**Fig. 3.5 A**). When triplicates of these experiments were analyzed and averaged by densitometry, they revealed an interesting pattern (**Fig. 3.5 B**). As expected, the R174D/R175D mutation significantly abrogated RNA binding, as it had for NTE binding. This decrease in binding was not as extreme as seen for NTE binding but was still on the order of two-fold. The K214D

mutation had no significant change in RNA binding. While this mutation did lead to a detectable change in NTE binding, it was still minor and so this effect is similar between the two.

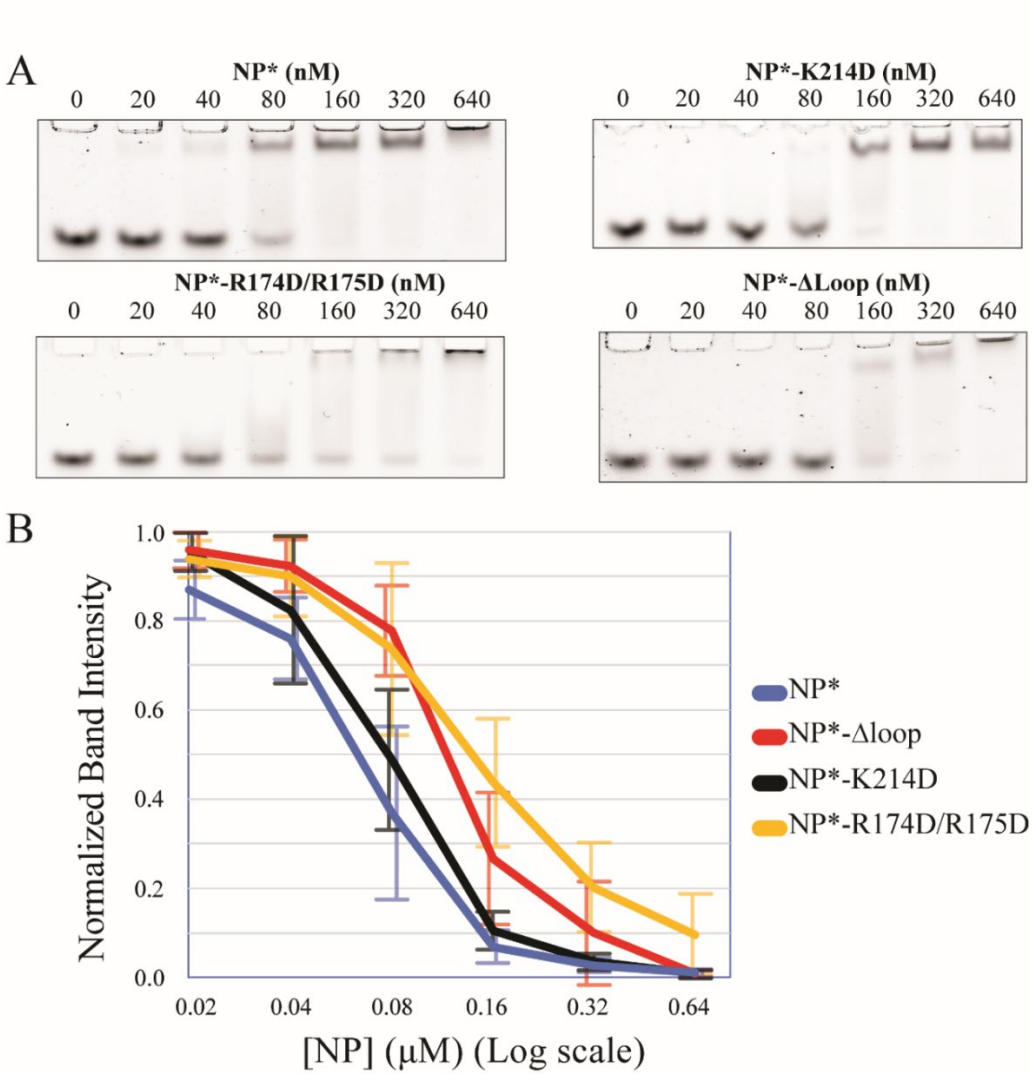

Figure 3.5 – NP* Loop Deletion and R174/R175D Mutations Abrogate RNA Binding

A) Native EMSA of 100 nM Alexa Fluor 488-labeled RNA binding to increasing amounts of NP*, NP*-K214D, NP*-Δloop or NP*-R174D/R175D, visualized by fluorescence. The lower band is unbound RNA, for NP*, NP*-K214D, and NP*-Δloop, an upper band of NP RNA complex forms as well. **B)** Quantification of RNA binding to all mutants tested from the native EMSAs presented in A). Bars represent standard deviation of band intensity across three images for each mutant. Panel A was adapted from Morris *et al*, 2020[3].

What was of some interest is that the ΔLoop mutant showed a notably reduced RNA binding profile, whereas there was no effect of this mutant on NTE binding. While this is not an extreme difference, the difference the loop plays in RNA binding vs NTE binding is significant, but without a structure or measurements of binding across the whole protein through scanning mutagenesis, it is difficult to interpret why this is the case. At any rate, the most important residues for NTE binding, R174 and R175, were also the most important for RNA binding.

A previous study by another group did scanning mutagenesis on NP and observed the effect on viral fitness as measured by viral titer in infected cells. They found that a mutation of R175A on a WT NP background dropped the titer more than half, while an R214A (K214 in PR8) produced only a minor decrease in titer, while K87 was not measured[209]. Our results are in line with the previous literature, as NP*-R174D/R175D's abrogation of RNA and NTE binding would be expected to severely impact viral replication, while our finding of the relative triviality of the K214D mutation is matched by their finding of its dispensability in vivo. Taken together, the binding data from the native EMSAs of RNA and FITC-UAP56-NTE$_{1\text{-}19}$ demonstrate that the binding site, as well as the binding mode towards NP, are very close and rely on a set of common residues in the center of the charged cleft.

3.3.5 – RNA and UAP56-NTE Compete for Binding to NP

We had found that UAP56-NTE and RNA occupy the same binding site in the basic cleft of NP, or at least an overlapping area, and that both are electrostatic interactions. The obvious implication of this is that RNA and NP should be able to compete with each other for binding to

NP. The test for this was to see if RNA could be used to disrupt an NTE-NP* complex. For this, FITC-UAP56-NTE$_{1-19}$ was pre-incubated with a large molar excess of NP* which was well above the dissociation constant, such that essentially all of the FITC-UAP56-NTE$_{1-19}$ was bound. Being bound in complex as such, it appeared at the top of the well when run on native-PAGE. When titrated amounts of RNA were added to this pre-formed complex, the RNA was capable of dislodging FITC-UAP56-NTE$_{1-19}$ so that it ran freely (**Fig. 3.6 A**). Of note, because of the large molar excess of NP*, at low concentrations there is a large pool of unbound NP* for the RNA to bind to without the need to compete. The tipping point seems to occur between 1 μM and 1.2 μM of RNA, where RNA can compete off the NTE. It is interesting to note how sharply the binding pivots between NTE and RNA. It is speculative to say, but where the NTE was seen to have an apparent cooperativity of binding, this may be true for RNA as a competitor, with RNA binding acting cooperatively to push NTE off NP*. What can be said conclusively though, is that RNA and UAP56 are competitors and that their binding is mutually exclusive.

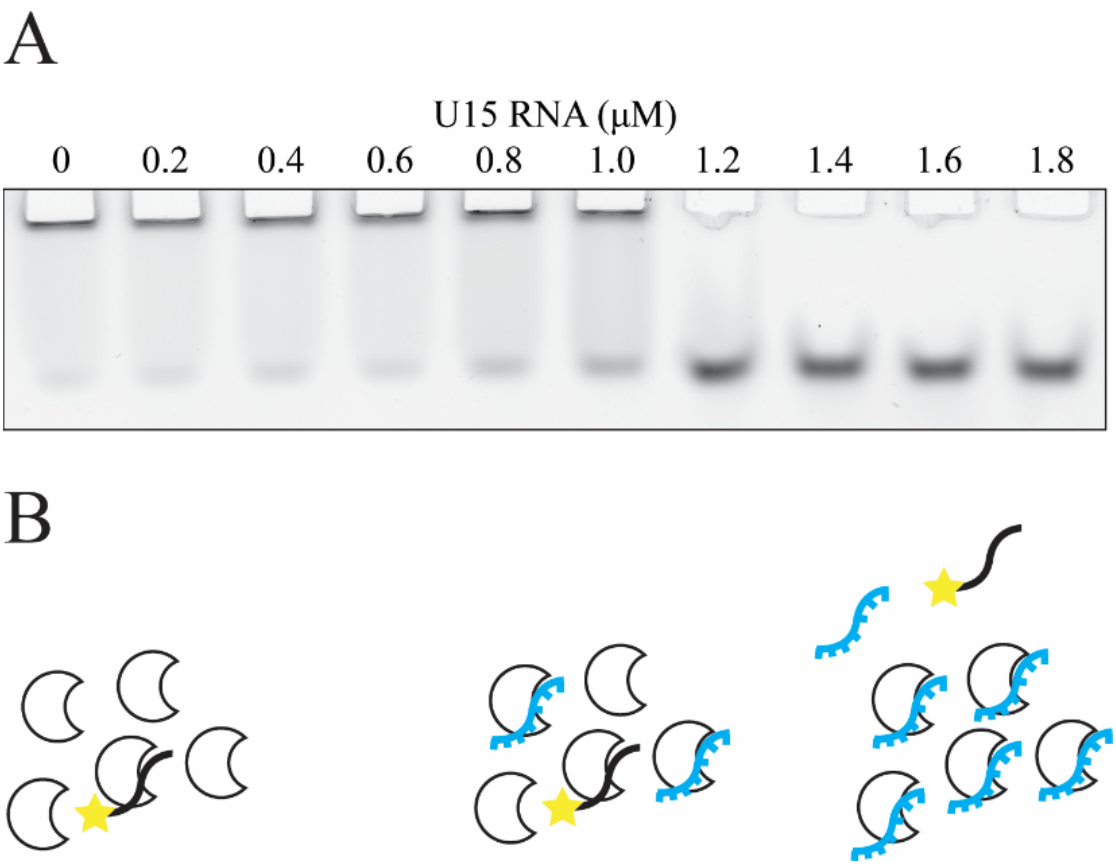

Figure 3.6 – RNA and UAP56-NTE Compete for Binding to NP

A) A native EMSA in which FITC-UAP56-NTE$_{1-19}$ is displaced from a molar excess of NP* by addition of increasing amounts RNA, visualized by fluorescence. **B)** Cartoon model of NP UAP56-NTE$_{1-19}$ competition with RNA. Panel A was adapted from Morris *et al*, 2020[3].

3.4 – Discussion

Where the first part of this study focused on the broader biochemistry of the UAP56-NP interaction, the goal of the latter end was of a structural view that built on the knowledge acquired up to that point. Our aim was a view of the UAP56-NP interaction at an atomic level. In this regard, we were mostly successful, although using a different methodology than we started with. There remain some open question, though

The first approach of X-ray crystallography seemed the most reasonable given the lab's expertise and the desire for a high-resolution view of the complex. Since WSN NP had been crystallized by several different groups with different ligands[153,154,192,206–208], we reasoned that co-crystallization with a small peptide should be tenable. It was encouraging that we were able to produce diffracting crystals with relative ease. The lack of NTE peptide density in the WT dataset after the soak was disappointing, though. We initially thought the issue was one of resolution, as our dataset with WSN WT only diffracted to 4 Å. When we were unable to see any density for a peptide with either the WSN R416A, or with the WSN Δ402-429 when soaked with either the UAP56-NTE$_{1-38}$ or FITC-UAP56-NTE$_{1-19}$, it became obvious that crystallography was no longer viable. At 2.8 Å, if the peptide were soaked in and occupied a well-ordered position, it would have been discernable if it were there. We still knew that the NTE was binding to NP, so why couldn't any electron density be seen? There are several factors that could account for this.

As noted before, the physical appearance of the crystals changed with the soak of FITC-UAP56-NTE$_{1-19}$, providing a way to know the fluorescently labeled peptide was incorporated. It might be possible that the color change seen here was merely the result of adsorption of the NTE to the surface of the crystals. This might explain why the optical properties of the crystal faces changed upon soaking. Given that crystals also became more brittle upon soaking suggests though that the peptide soak was complete and that it diffused all the way through. It's more likely that the NTE peptide was able to completely soak into the crystals and that its absence in the solved structure is the result of either occupying solvent channels or of a broad open binding surface that can accommodate a broad range of conformations and positions of the peptide such that it is disordered and will not produce diffraction. Solvent channels exist when the unit cell contains large region occupied by water which are contiguous with neighboring cells and form

passageways for small molecules to diffuse into the crystal by displacing water without ever binding or interacting with the protein specifically. Visual inspection of the unit cell packing for the published structures of the three constructs which we utilized shows that this is a possible explanation in the case of WSN R416A or for WSN Δ402-429, which appear porous at the nanometer scale, where a peptide could slip in. For WSN WT however, the solvent fraction is quite low, and the asymmetric units are quite densely packed, and this seems an unlikely explanation. The most plausible reason for the lack of peptide density is that while the soak was successful, and the NTE peptide was bound, that the binding was highly accommodating to a variety of peptide conformations.

While a crystal structure would have been ideal to provide an exact picture of the molecular interface, similar information, albeit in less detail, was obtained by the combination of XL-MS and mutagenesis. The identification of K87 and K214 on the opposite edges of the RNA binding cleft gave support to the hypothesis that this is where NTE was binding, which was in line with the expectation that the interaction should be electrostatically driven. When these peripheral lysines were mutated however, little abrogation was seen in NTE binding. By contrast, the arginines in the center of the cleft at R174 and R175, which are located between K87 and K214, are critical for NTE binding as demonstrated by our native EMSA titration experiments. This demonstrated that the UAP56-NTE NP interaction is indeed occurring in the RNA binding cleft, and is electrostatically driven as hypothesized. Native EMSA titrations with fluorescent RNA as the substrate showed that R174 and R175 are also important in RNA binding. This supported the idea that UAP56-NTE and RNA have at least a partially overlapping binding site and should be able to compete with one another for binding to NP in a mutually exclusive manner. This was

confirmed by a simple competition experiment, once again by native EMSA, which showed the ability of RNA to displace FITC-UAP56-NTE$_{1-19}$ from a complex with NP.

In light of the XL-MS results and the differential effects of mutagenesis on both UAP56-NTE and RNA binding, the inability to see NTE peptide density in crystals which otherwise diffracted well makes more sense. The fact that the terminal glutamate of UAP56-NTE$_{1-19}$ was seen to cross link to both K214 and K87, which are on opposite sides of the RNA binding cleft, around 30 Å from each other could only be possible if a variety of conformational states of the peptide could be accommodated, but with the crux of the interaction occurring at in the center of the cleft at R174 and R175. There are insufficient data to be able to extrapolate the occupancy of the NTE in different states along the surface of the RNA binding cleft. However, given the distance between the residues that were detected by XL-MS, and the fact that there are many positively charged residues spread out over a wide surface in the cleft, with many negatively charged residues on the flexible NTE, it seems hard to imagine that there is one bound formation which would be so highly preferred over the rest. So, a rough model for the interaction is more analogous to magnets on a refrigerator than to Cinderella's glass slipper, but it usually only the latter type of interaction which will diffract in x-ray crystallography. This might also explain why NP has never been crystallized with RNA bound, as the lack of a rigid fit with the substrate is not as amenable to crystallization.

Single-stranded RNA is highly flexible and can take on many different secondary and tertiary structures, and NP binds it with no notable sequence specificity. In this context, an interaction driven by the avidity of many electrostatic interactions as opposed to steric selectivity makes sense. UAP56-NTE and RNA have commonality in that both are densely negatively charged and flexible, but they are sterically different. If a precise geometry of specific salt bridges was

required for RNA binding, then it would be very unlikely that UAP56-NTE, which shares at least a partially overlapping binding surface on NP, could accommodate that exact set of constraints in such a way as for the binding of UAP56-NTE and RNA to be mutually exclusive. A lock and key arrangement could not accommodate binding by substrates with considerably different shapes. This may also explain how the NP*-ΔLoop mutant had no effect on NTE binding but produced a significant abrogation of RNA binding. Although the central arginines are the most important contact point for both substrates, differences in shape and distribution of charge might make different auxiliary contact points more important in NTE binding versus RNA binding given the wealth of charged surface residues the substrate could alternately bind to.

CHAPTER 4

CONCLUSIONS AND FUTURE DIRECTIONS

4.1 – Conclusions

The eukaryotic DEAD-box helicase UAP56 is a protein which has a role in the cell's processing of nascent mRNPs as part of the TREX complex, and is required for proper export of most mRNPs into the cytoplasm[2,31,66]. Several studies have also demonstrated that it is a critical component in the lifecycle of the Influenza A virus. In this role, it was shown to be critical in the replication of viral ribonucleoproteins, which are the packaged unit of IAV's single stranded RNA genome[173,179,187,190]. It acts to prevent the abortive replication of cRNAs and vRNAs by the viral polymerase by binding to influenza NP, which is the RNA binding protein which coats the vRNP, and recruits it to the nascent vRNA as it emerges from the viral polymerase. It was thought that this occurred through the C-terminal RecA domain of UAP56, through interaction with NP trimers which were not RNA bound.

This was the extent of the knowledge of the UAP56 interaction with NP when we began this project. Our work has led to a few notable refinements of the basic understanding of this, though a great amount still remains unknown. Our aim when starting this project was to uncover the critical regions of UAP56 and NP that constituted the interface between the two, and then building off of that, to find the specific residues that formed the interaction, ideally through an atomic resolution structure. Then finally, based on that information, we would verify their impact on binding in vitro.

The most critical findings of our studies can be summarized as follows. First, the highly negatively charged and disordered N-terminal extension of UAP56, specifically the first 19 residues are sufficient to see binding of NP at a strength comparable to that of the full-length protein. While the NTE is the critical interacting element, a weaker binding can be seen between NP and the two RecA domains of UAP56. Second, this interaction can be seen across strains, both PR8 and WSN, as well as for monomeric mutants of NP. Binding is considerably more robust for WT NP, however. Thirdly, the UAP56 NP interaction is at the center of the highly positively charged RNA binding cleft of NP, and is formed by an electrostatic interaction, with the central arginines R174 and R175 being critical. Finally, we show that because of occupying the same binding interface on NP, binding of RNA to NP is competitive to UAP56 binding.

From these findings, we put forward a very basic model of the UAP56 NP interaction and its relation to viral transcription which builds on the existing view from the literature. First, UAP56, with its NTE disengaged from the core, will bind to RNA-free NP, either as a monomer or oligomer. It does so through an electrostatic interaction between the NTE and the RNA binding cleft of NP (**Fig. 4.1 A**). The UAP56 NP complex then arrives at nascent vRNA or cRNA as it emerges proximal to the viral polymerase (**Fig. 4.1 B**). UAP56, likely by binding to this naked RNA, then deposits NP onto the growing vRNP or cRNA, and in doing so, prevents abortive replication. UAP56 then disengages and leaves for this cycle to repeat until replication is

completed (**Fig. 4.1 C**). There are many areas of this model which are unclear and require further study, as is discussed below.

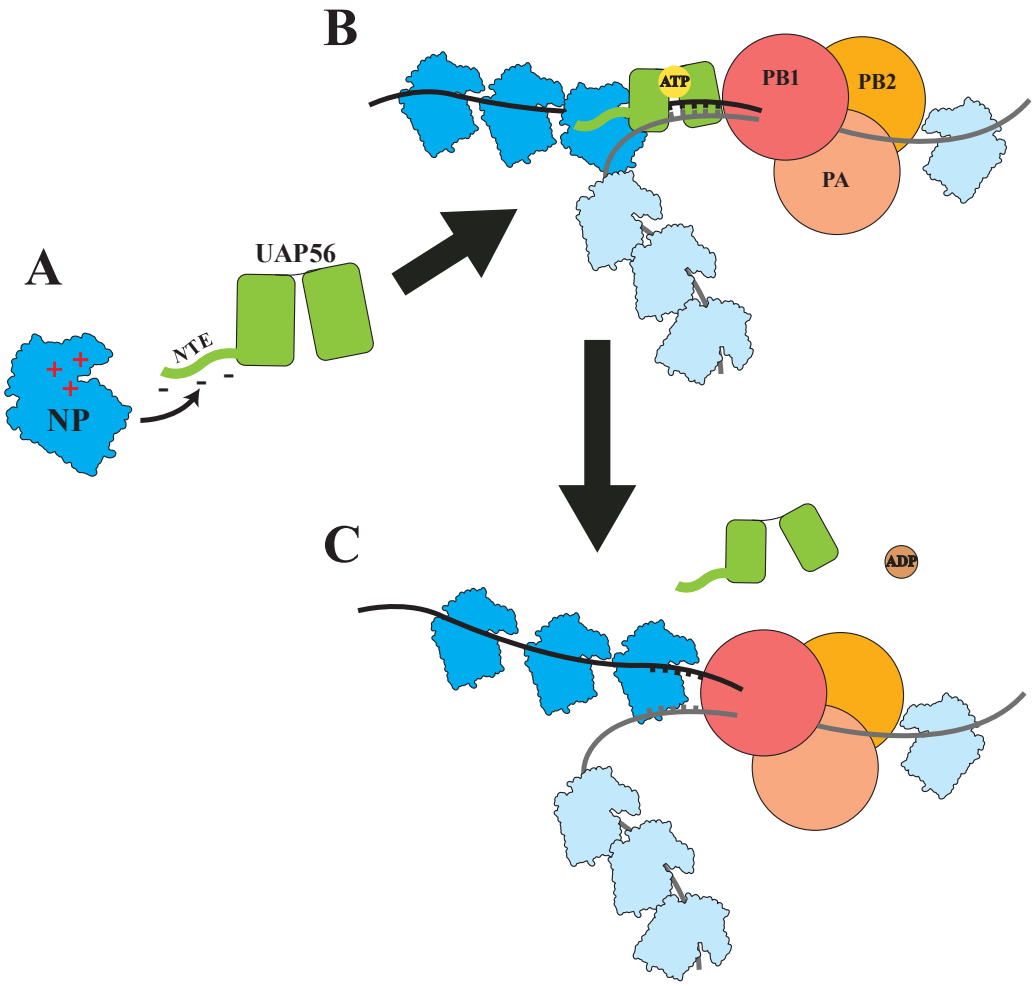

Figure 4.1 – A Possible Model of UAP56 NP Interaction

A general schematic of the mode of action of UAP56 on NP during viral replication, based on our findings. Nascent vRNA is shown as a black strand, with template vRNA in grey and NP associated with the template in light blue. **A)** NP not bound to RNA binds with the UAP56 exposed NTE to its RNA binding cleft by electrostatics. **B)** UAP56 brings the NP to the nascent and exposed vRNA as it emerges from the viral polymerase **C)** UAP56 deposits NP onto the RNA and leaves the complex. Presumably, this requires the ATPase and helicase activity of UAP56 to transfer RNA to NP and disengage and possibly to unwind nascent vRNA which has reannealed to the template.

Our results refine the understanding of the UAP56 NP interaction which existed before our study. We were able to show that UAP56 is capable of interacting even with the monomeric mutant NP*. Recent work by the Nagata group, who had initially characterized the interaction, had indicated that UAP56 interacts only with NP in a trimeric form[190]. While plausible evidence was provided to support this, it could be interpreted multiple ways and is not definitive. Notably, we saw evidence of UAP56 interaction with NP*, which would not be predicted from their model. It is not wholly contradictory to their results though, as binding to multimeric NP is certainly considerably more robust, as indicated by the higher salt tolerance of that interaction during pulldowns, and the inability of GST-UAP56-Core to pulldown NP*. Additionally, while UAP56 NP* occurs, its conditions might not be physiological. Our pulldown method is also much gentler than their method of co-migration of size exclusion chromatography. Curiously, this paper also showed data which demonstrated that UAP56 promoted or accelerated the multimerization of NP in vitro. It is hard to imagine that UAP56 could do this if it is not capable of binding to the monomeric NP population, unless it is only increasing the size of existing multimers. This however, does make sense in light of our finding that interaction with the monomer is possible. Conversely, their finding of induced multimerization sheds light on a curious trend in our data. Our MST and fluorescence polarization data did not fit a 1-to-1 binding model between UAP56 and NP*, instead fitting with a Hill coefficient of over 2, indicating an apparent cooperativity in binding. In light of their results, it might be possible that NP* binding to the NTE promotes the formation of multimeric NP to some extent despite its otherwise inability to do so. Perhaps the NTE is forming a linker between NP* monomers at higher concentrations through binding to part of the RNA binding site of one at one end and part of the

other site on the other end. This might explain the otherwise puzzling apparent cooperativity of biding. Our results also corroborate the findings of this paper in demonstrating competition between NP and RNA for UAP56 binding, an interaction which they had also found to be mutually exclusive by in vitro methods.

The first paper from the Nagata group on the NP UAP56 interaction reported that the minimal interacting region of UAP56 lied in a fragment of amino acid residues 248 – 428 encompassing the C-terminal RecA domain, which they identified by co-immunoprecipitation using tagged NP of scanning deletions of UAP56[187]. Our results agree with this. We are able to see this interaction as well when we were able to pull down NP with GST-UAP56-Core, which includes this region. It was, however, considerably less robust in binding compared to the full length. In contrast, we identified the N-terminal element as being the strongest binding region, which did not show up in their deletion scan. It is not immediately clear why they were unable to see this, but degradation is one possibility, since the NTE is exposed. They also reported that the minimal binding region of NP consisted of the first 20 residues when they reversed the design of their co-immunoprecipitation experiment and did scanning deletions of NP. While we did not test this, given our discovery of the UAP56 NTE as the minimal interacting region, it seems unlikely that two small peptides could form a functional interface and furthermore, residues 1-20 are not within the NP RNAS binding cleft.

Curiously, this paper also reported that the ATPase activity of UAP56 is not required for its activity as an NP recruitment factor and viral antitermination factor[179]. As ATP hydrolysis is necessary for the helicase activity of UAP56, this raises the question of whether that helicase activity is necessary at all, and if not, how UAP56 is then released from the vRNA. While our results shed light on the manner of UAP56 NP interaction, the overall mechanism of how UAP56

is acting as an antitermination factor is essentially a black box where all that is really known is that it is necessary. It is certainly intuitive to imagine that UAP56 is performing this role by its helicase activity. The viral polymerase lacks its own helicase activity, and some form of active remodeling factor has to be required to dislodge NP to allow the helicase to thread the template through when transcribing or replicating. The template and product would also be prone to reannealing without a helicase, unless exiting strands were bound immediately as they emerged.

4.2 – Open Questions

4.2.1 – In Vivo Significance

Our findings open up a variety of new questions and leave open some existing ones. Many of these questions this lab is not suited to tackle but should at some point be addressed for a more complete understanding in both the fields of mRNA export and influenza biology.

At one point in our studies, we had seriously considered testing our findings in vivo in influenza infected cells through a collaborator. While previous studies have done this with siRNA mediated knockdown of UAP56 looking at the impact on viral replication[179,210], we were hesitant as mutation or knockdown of UAP56 was likely to be quite toxic to the host cell and thus difficult to separate changes specific to viral replication from the effects of a host which is generally sick to begin with. However, at some point our findings will need to be shown to have relevance in cells. It would be a gamble to try, but one could try by demonstrating that transient knockdown of endogenous UAP56 could produce attenuated viral titer independent of any effect on the expression of interferon stimulated genes, while also producing minimal toxicity in uninfected control knockdown cells. If this abrogation of viral titer could not be rescued by

simultaneous transient expression of a UAP56-Core, this would show that UAP56-NTE is critical not just to in vitro binding, but in vivo for viral replication.

A mini-genome replication system also exists that allows for monitoring effects on transcription and replication of individual influenza genome segments by plasmid expression rather than infection and has been used many times to study mutations that could not allow functional virions to be produced[165,211,212]. Such a system could also be used here to study mutation of NP in mammalian cells. The description of the above experiments brushes over many confounding factors that would make them difficult to try, though. At some point however, someone should try to look at the UAP56 NP interaction in detail in a physiological environment.

4.2.2 – Further Mutational Studies

Our studies were limited in looking at primarily interactions between UAP56 and NP*. NP* is useful for its superior solubility and uniform behavior across a variety of conditions. It also allows the disentanglement of the effects of NP oligomerization on the UAP56 affinity for NP and so allowed us a cleaner, simpler system to work with. An obligate monomer is not physiological, however. In vivo, monomeric mutants of NP produce inviable virus. The most critical residues we identified on NP*, R174 and R175, should be examined as mutants on the WT PR8 NP background, to demonstrate the physiological relevance and show that the residues critical for NTE interaction with NP* are the same for NP in an oligomer. Indeed, likely most of the further studies examining the interaction should be done with PR8 WT NP when possible.

While we have established that the center of the RNA binding cleft is essential for the UAP56 NP interaction, our data is too sparse to provide a complete view of the occupancy of the NTE

peptide within the cleft. The preferred orientation of the NTE within the cleft, if there is one, is not known. Our pulldown data show that UAP56-core is capable of independent interaction with NP, in agreement with the original paper describing the interaction. The XL-MS was only done with the NTE however, so this interaction was not probed for the specific residues making contact in the UAP56-Core. For a complete look at all of the individual contacts being made over the whole of the protein, the mass spectrometry should be expanded. Originally, we had planned to do hydrogen-deuterium exchange mass spectrometry instead of crosslinking. This method has the advantage of not being limited by a specific chemistry needed for crosslinking and can show the solvent accessibility of all residues to show which are being buried in an interface. So, in order to obtain a complete picture at a residue-specific level over the entire length of both proteins, hydrogen-deuterium exchange mass spectrometry should be performed using UAP56-FL with NP or NP*.

Additionally, it is unlikely that the central arginines alone are the main determinant of NP-UAP56 interaction. R174 and R175 are not the only positive charges in the center of the cleft. Other likely candidates include R150, R152, R156, R195, R199, and R221 based on proximity, and that they form a path of very high positive charge density along the center of the cleft (**Fig. 4.2**). Certainly, exhaustive testing of all positive residues in the RNA binding cleft could be labor intensive. In this regard, the findings from hydrogen-deuterium exchange mass spectrometry would be informative in narrowing the next sites for mutagenesis, and since we have demonstrated that the NTE and RNA bind by the same mechanism, future mutagenesis could also be guided by previous mutagenic studies examining the residues critical for RNA binding.

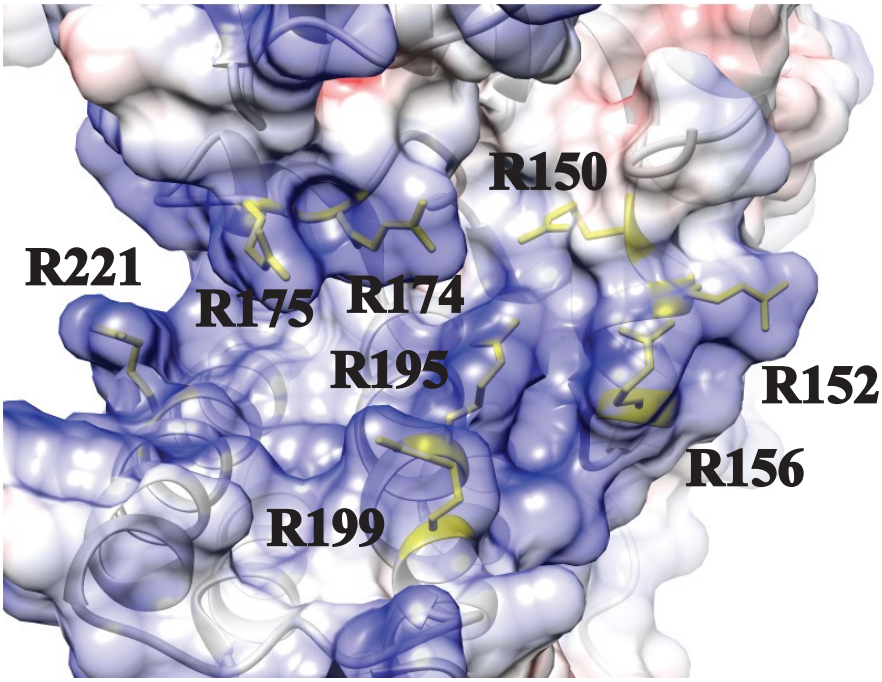

Figure 4.2 – Possible NP Residues to Examine for UAP56 Interaction by Mutagenesis

Structure of WSN NP R416A structure (PDB: 3ZDP)[154] showing conserved residues in yellow which could outline a more complete NTE binding in the center of the RNA binding cleft. This is overlaid with an electrostatic surface map, on which blue indicates positive charge, and red negative. Figure was generated with UCSF Chimera[69].

4.2.3 – The Role of the TREX Complex

The effect of NP binding on UAP56 was unexplored in this study, and this should be examined in future studies. The NTE of UAP56 has a known function in regulating the ATPase activity of the protein[66], which is a feature conserved with several other DEAD/DECD-box helicases including DDX25[76], and DDX19[74]. If the NTE is sequestered by binding to NP, then it is presumably unavailable to regulate ATPase activity this way. So, NP should then act to suppress the ATPase activity of UAP56. We tried to measure this in vitro but encountered a few problems. In light of later results that the NTE and RNA compete for access to NP, we can see that one problem with this assay is that UAP56 requires RNA for ATPase hydrolysis. So, having an

excess of NP present acts to sequester the RNA from UAP56, while conversely having an excess of RNA sequesters NP from binding. This makes it difficult to establish the complex in the desired condition to test the effect. The other issue is that UAP56 has an intrinsically very slow rate of ATP hydrolysis, making the signal to noise of the assay quite small. Accurately measuring the ATPase activity assay for UAP56 would likely require scintillation methods but might provide insight into how UAP56 is performing its role as a recruitment factor to vRNPs. Performing these experiments is all the more relevant given the claim that the ATPase activity of UAP56 is unnecessary[179].

One other question that the field must consider is the role of UAP56 in the TREX complex. The cellular functions of UAP56 do not act in isolation, but as a concerted series of binding events involving the EJC, THO complex and ALY. When Influenza is conscripting UAP56, is it doing so with UAP56 free in solution or at some point in the assembly of the TREX complex? There seems reason to speculate that the latter might be true. The influenza polymerase itself will associate with the Pol II CTD to place it in proximity to cap snatch during viral transcription[213,214]. The physical linkage of the components of mRNA transcription and processing are vital for the host, and it would make sense for the virus to exploit this. Could UAP56 be recruited while the viral polymerase is in complex with the Pol II CTD and the TREX complex is physically linked to it through the host mRNP? This would certainly provide the virus with easy access. If UAP56 is recruited from the host mRNP then would its association with other TREX complex members enhance or abrogate its role in vRNP assembly? The former question might be addressed with co-localization experiments such as a proximity ligation assay in cells, while the latter might be examined by the in vitro viral replication assays, which demonstrated UAP56's role in viral anti-termination, with the addition of THO components or

ALY. Of course, all of this is all extremely speculative, and examining any of these questions would be quite an ambitious project in its own right. Eventually though, a complete view of viral replication needs to consider the larger context of all of the host's mRNA processing machinery.

In summary, there is a large amount we still have yet to discover about the fine points of the machinations underpinning UAP56 function in viral replication. Certainly, no one series of experiments will be able to provide a complete picture and elucidating this is a long-term goal which will require the devoted efforts of at least several different labs. Our findings however, provide novel and significant insight into this interaction and open up exciting new avenues of investigation.

REFERENCES

1. Grüter, P. *et al.* TAP, the human homolog of Mex67p, mediates CTE-dependent RNA export from the nucleus. *Mol. Cell* (1998) doi:10.1016/S1097-2765(00)80065-9.

2. Xie, Y. & Ren, Y. Mechanisms of nuclear mRNA export: A structural perspective. *Traffic* **20**, 829–840 (2019).

3. Morris, A. K. *et al.* Cellular mRNA export factor UAP56 recognizes nucleic acid binding site of influenza virus NP protein. *Biochem. Biophys. Res. Commun.* (2020) doi:10.1016/j.bbrc.2020.02.059.

4. Coller, J. & Parker, R. Eukaryotic mRNA Decapping. *Annu. Rev. Biochem.* (2004) doi:10.1146/annurev.biochem.73.011303.074032.

5. Gonatopoulos-Pournatzis, T. & Cowling, V. H. Cap-binding complex (CBC). *Biochemical Journal* (2014) doi:10.1042/BJ20131214.

6. Gonatopoulos-Pournatzis, T., Dunn, S., Bounds, R. & Cowling, V. H. RAM/Fam103a1 is required for mRNA cap methylation. *Mol. Cell* **44**, 585–96 (2011).

7. Geuens, T., Bouhy, D. & Timmerman, V. The hnRNP family: insights into their role in health and disease. *Human Genetics* (2016) doi:10.1007/s00439-016-1683-5.

8. Filipowicz, W. & Pogači, V. Biogenesis of small nucleolar ribonucleoproteins. *Current Opinion in Cell Biology* (2002) doi:10.1016/S0955-0674(02)00334-4.

9. Rearick, D. *et al.* Critical association of ncRNA with introns. *Nucleic Acids Res.* (2011)

doi:10.1093/nar/gkq1080.

10. Shi, Y. Mechanistic insights into precursor messenger RNA splicing by the spliceosome. *Nature Reviews Molecular Cell Biology* (2017) doi:10.1038/nrm.2017.86.

11. Nelson, D. L. & Cox, M. M. RNA Metabolism. in *Lehninger Principles of Biochemistry* (ed. Ahr, K.) 1036–1039 (W. H. Freeman and Company, 2008).

12. Le Hir, H., Izaurralde, E., Maquat, L. E. & Moore, M. J. The spliceosome deposits multiple proteins 20-24 nucleotides upstream of mRNA exon-exon junctions. *EMBO J.* (2000) doi:10.1093/emboj/19.24.6860.

13. Hir, H. Le, Saulière, J. & Wang, Z. The exon junction complex as a node of post-transcriptional networks. *Nature Reviews Molecular Cell Biology* (2016) doi:10.1038/nrm.2015.7.

14. Luo, M. J. & Reed, R. Splicing is required for rapid and efficient mRNA export in metazoans. *Proc. Natl. Acad. Sci. U. S. A.* (1999) doi:10.1073/pnas.96.26.14937.

15. Masuda, S. *et al.* Recruitment of the human TREX complex to mRNA during splicing. *Genes Dev.* (2005) doi:10.1101/gad.1302205.

16. Hug, N., Longman, D. & Cáceres, J. F. Mechanism and regulation of the nonsense-mediated decay pathway. *Nucleic Acids Research* (2015) doi:10.1093/nar/gkw010.

17. Proudfoot, N. J. & Brownlee, G. G. 3′ Non-coding region sequences in eukaryotic messenger RNA. *Nature* (1976) doi:10.1038/263211a0.

18. Kumar, A., Clerici, M., Muckenfuss, L. M., Passmore, L. A. & Jinek, M. Mechanistic insights into mRNA 3′-end processing. *Current Opinion in Structural Biology* (2019)

doi:10.1016/j.sbi.2019.08.001.

19. Mandel, C. R. *et al.* Polyadenylation factor CPSF-73 is the pre-mRNA 3′-end-processing endonuclease. *Nature* (2006) doi:10.1038/nature05363.

20. Proudfoot, N. New perspectives on connecting messenger RNA 3′ end formation to transcription. *Current Opinion in Cell Biology* (2004) doi:10.1016/j.ceb.2004.03.007.

21. Mandel, C. R., Bai, Y. & Tong, L. Protein factors in pre-mRNA 3′-end processing. *Cellular and Molecular Life Sciences* (2008) doi:10.1007/s00018-007-7474-3.

22. Chavez, S. A protein complex containing Tho2, Hpr1, Mft1 and a novel protein, Thp2, connects transcription elongation with mitotic recombination in Saccharomyces cerevisiae. *EMBO J.* (2000) doi:10.1093/emboj/19.21.5824.

23. Strabetaer, K. Yra1p, a conserved nuclear RNA-binding protein, interacts directly with Mex67p and is required for mRNA export. *EMBO J.* (2000) doi:10.1093/emboj/19.3.410.

24. Sträßer, K. & Hurt, E. Splicing factor Sub2p is required for nuclear mRNA export through its interaction with Yra1p. *Nature* (2001) doi:10.1038/35098113.

25. Stäßer, K. *et al.* TREX is a conserved complex coupling transcription with messenger RNA export. *Nature* (2002) doi:10.1038/nature746.

26. Gross, S. & Moore, C. Five subunits are required for reconstitution of the cleavage and polyadenylation activities of Saccharomyces cerevisiae cleavage factor I. *Proc. Natl. Acad. Sci. U. S. A.* (2001) doi:10.1073/pnas.101046598.

27. Johnson, S. A., Cubberley, G. & Bentley, D. L. Cotranscriptional Recruitment of the mRNA Export Factor Yra1 by Direct Interaction with the 3′ End Processing Factor Pcf11.

Mol. Cell (2009) doi:10.1016/j.molcel.2008.12.007.

28. Johnson, S. A., Kim, H., Erickson, B. & Bentley, D. L. The export factor Yra1 modulates mRNA 3′ end processing. *Nat. Struct. Mol. Biol.* (2010) doi:10.1038/nsmb.2126.

29. Shi, M. *et al.* ALYREF mainly binds to the 5′ and the 3′ regions of the mRNA in vivo. *Nucleic Acids Res.* (2017) doi:10.1093/nar/gkx597.

30. Morris, K. J. & Corbett, A. H. The polyadenosine RNA-binding protein ZC3H14 interacts with the THO complex and coordinately regulates the processing of neuronal transcripts. *Nucleic Acids Res.* (2018) doi:10.1093/nar/gky446.

31. Carmody, S. R. & Wente, S. R. mRNA nuclear export at a glance. *J. Cell Sci.* **122**, 1933–1937 (2009).

32. Ren, Y., Schmiege, P. & Blobel, G. Structural and biochemical analyses of the DEAD-box ATPase Sub2 in association with THO or Yra1. *Elife* (2017) doi:10.7554/eLife.20070.

33. Meinel, D. M. *et al.* Recruitment of TREX to the Transcription Machinery by Its Direct Binding to the Phospho-CTD of RNA Polymerase II. *PLoS Genet.* (2013) doi:10.1371/journal.pgen.1003914.

34. Zenklusen, D., Vinciguerra, P., Wyss, J.-C. & Stutz, F. Stable mRNP Formation and Export Require Cotranscriptional Recruitment of the mRNA Export Factors Yra1p and Sub2p by Hpr1p. *Mol. Cell. Biol.* (2002) doi:10.1128/mcb.22.23.8241-8253.2002.

35. Chi, B. *et al.* Aly and THO are required for assembly of the human TREX complex and association of TREX components with the spliced mRNA. *Nucleic Acids Res.* **41**, 1294–306 (2013).

36. Zhou, Z. *et al.* The protein Aly links pre-messenger-RNA splicing to nuclear export in metazoans. *Nature* (2000) doi:10.1038/35030160.

37. Viphakone, N. *et al.* TREX exposes the RNA-binding domain of Nxf1 to enable mRNA export. *Nat. Commun.* (2012) doi:10.1038/ncomms2005.

38. Longman, D., Johnstone, I. L. & Cáceres, J. F. The REF/Aly proteins are dispensable for mRNA export and development in Caenorhabditis elegans. *RNA* **9**, 881–891 (2003).

39. Gatfield, D. & Izaurralde, E. REF1/Aly and the additional exon junction complex proteins are dispensable for nuclear mRNA export. *J. Cell Biol.* **159**, 579–588 (2002).

40. Katahira, J., Inoue, H., Hurt, E. & Yoneda, Y. Adaptor Aly and co-adaptor Thoc5 function in the Tap-p15-mediated nuclear export of HSP70 mRNA. *EMBO J.* (2009) doi:10.1038/emboj.2009.5.

41. Katahira, J., Dimitrova, L., Imai, Y. & Hurt, E. NTF2-like domain of Tap plays a critical role in cargo mRNA recognition and export. *Nucleic Acids Res.* (2015) doi:10.1093/nar/gkv039.

42. Taniguchi, I. & Ohno, M. ATP-Dependent Recruitment of Export Factor Aly/REF onto Intronless mRNAs by RNA Helicase UAP56. *Mol. Cell. Biol.* **28**, 601–608 (2008).

43. Chang, C. Te *et al.* Chtop is a component of the dynamic TREX mRNA export complex. *EMBO J.* (2013) doi:10.1038/emboj.2012.342.

44. Iglesias, N. *et al.* Ubiquitin-mediated mRNP dynamics and surveillance prior to budding yeast mRNA export. *Genes Dev.* (2010) doi:10.1101/gad.583310.

45. Stutz, F. *et al.* REF, an evolutionarily conserved family of hnRNP-like proteins, interacts

with TAP/Mex67p and participates in mRNA nuclear export. *RNA* **6**, 638–650 (2000).

46. Golovanov, A. P., Hautbergue, G. M., Tintaru, A. M., Lian, L.-Y. & Wilson, S. A. The solution structure of REF2-I reveals interdomain interactions and regions involved in binding mRNA export factors and RNA. *RNA* **12**, 1933–48 (2006).

47. Rodrigues, J. P. *et al.* REF proteins mediate the export of spliced and unspliced mRNAs from the nucleus. *Proc. Natl. Acad. Sci. U. S. A.* (2001) doi:10.1073/pnas.98.3.1030.

48. Hautbergue, G. M., Hung, M. L., Golovanov, A. P., Lian, L. Y. & Wilson, S. A. Mutually exclusive interactions drive handover of mRNA from export adaptors to TAP. *Proc. Natl. Acad. Sci. U. S. A.* (2008) doi:10.1073/pnas.0709167105.

49. Huang, Y., Gattoni, R., Stévenin, J. & Steitz, J. A. SR splicing factors serve as adapter proteins for TAP-dependent mRNA export. *Mol. Cell* (2003) doi:10.1016/S1097-2765(03)00089-3.

50. Tuck, A. C. & Tollervey, D. A transcriptome-wide atlas of RNP composition reveals diverse classes of mRNAs and lncRNAs. *Cell* (2013) doi:10.1016/j.cell.2013.07.047.

51. Baejen, C. *et al.* Transcriptome Maps of mRNP Biogenesis Factors Define Pre-mRNA Recognition. *Mol. Cell* (2014) doi:10.1016/j.molcel.2014.08.005.

52. Hargous, Y. *et al.* Molecular basis of RNA recognition and TAP binding by the SR proteins SRp20 and 9G8. *EMBO J.* (2006) doi:10.1038/sj.emboj.7601385.

53. Blevins, M. B., Smith, A. M., Phillips, E. M. & Powers, M. A. Complex formation among the RNA export proteins Nup98, Rae1/Gle2, and TAP. *J. Biol. Chem.* **278**, 20979–20988 (2003).

54. Herold, A. *et al.* TAP (NXF1) Belongs to a Multigene Family of Putative RNA Export Factors with a Conserved Modular Architecture. *Mol. Cell. Biol.* (2000) doi:10.1128/mcb.20.23.8996-9008.2000.

55. Segref, A. *et al.* Mex67p, a novel factor for nuclear mRNA export. Binds to both poly(A)+ RNA and nuclear pores. *EMBO J.* (1997) doi:10.1093/emboj/16.11.3256.

56. Joong Kim, S. *et al.* Integrative structure and functional anatomy of a nuclear pore complex. *Nat. Publ. Gr.* (2018) doi:10.1038/nature26003.

57. Schmidt, H. B. & Görlich, D. Transport Selectivity of Nuclear Pores, Phase Separation, and Membraneless Organelles. *Trends in Biochemical Sciences* (2016) doi:10.1016/j.tibs.2015.11.001.

58. Mehlin, H., Daneholt, B. & Skoglund, U. Structural interaction between the nuclear pore complex and a specific translocating RNP particle. *J. Cell Biol.* (1995) doi:10.1083/jcb.129.5.1205.

59. Yang, W., Gelles, J. & Musser, S. M. Imaging of single-molecule translocation through nuclear pore complexes. *Proc. Natl. Acad. Sci. U. S. A.* (2004) doi:10.1073/pnas.0403675101.

60. Tran, E. J. & Wente, S. R. Dynamic Nuclear Pore Complexes: Life on the Edge. *Cell* (2006) doi:10.1016/j.cell.2006.05.027.

61. Tran, E. J., Zhou, Y., Corbett, A. H. & Wente, S. R. The DEAD-Box Protein Dbp5 Controls mRNA Export by Triggering Specific RNA:Protein Remodeling Events. *Mol. Cell* (2007) doi:10.1016/j.molcel.2007.09.019.

62. Weirich, C. S. *et al.* Activation of the DExD/H-box protein Dbp5 by the nuclear-pore protein Gle1 and its coactivator InsP6 is required for mRNA export. *Nat. Cell Biol.* **8**, 668–76 (2006).

63. Tseng, S. S. I. *et al.* Dbp5p, a cytosolic RNA helicase, is required for poly(A)+ RNA export. *EMBO J.* (1998) doi:10.1093/emboj/17.9.2651.

64. Rudolph, M. G. & Klostermeier, D. When core competence is not enough: Functional interplay of the DEAD-box helicase core with ancillary domains and auxiliary factors in RNA binding and unwinding. *Biological Chemistry* (2015) doi:10.1515/hsz-2014-0277.

65. Sengoku, T., Nureki, O., Nakamura, A., Kobayashi, S. & Yokoyama, S. Structural Basis for RNA Unwinding by the DEAD-Box Protein Drosophila Vasa. *Cell* (2006) doi:10.1016/j.cell.2006.01.054.

66. Shi, H., Cordin, O., Minder, C. M., Linder, P. & Xu, R.-M. Crystal structure of the human ATP-dependent splicing and export factor UAP56. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 17628–33 (2004).

67. Caruthers, J. M., Johnson, E. R. & McKay, D. B. Crystal structure of yeast initiation factor 4A, a DEAD-box RNA helicase. *Proc. Natl. Acad. Sci. U. S. A.* (2000) doi:10.1073/pnas.97.24.13080.

68. Linder, P. & Jankowsky, E. From unwinding to clamping ĝ€" the DEAD box RNA helicase family. *Nature Reviews Molecular Cell Biology* (2011) doi:10.1038/nrm3154.

69. Pettersen, E. F. *et al.* UCSF Chimera - A visualization system for exploratory research and analysis. *J. Comput. Chem.* (2004) doi:10.1002/jcc.20084.

70. Rudolph, M. G. & Klostermeier, D. The Thermus thermophilus DEAD box helicase Hera contains a modified RNA recognition motif domain loosely connected to the helicase core. *RNA* (2009) doi:10.1261/rna.1820009.

71. Diges, C. M. & Uhlenbeck, O. C. Escherichia coli DbpA is an RNA helicase that requires hairpin 92 of 23S rRNA. *EMBO J.* (2001) doi:10.1093/emboj/20.19.5503.

72. Del Campo, M. & Lambowitz, A. M. Structure of the Yeast DEAD Box Protein Mss116p Reveals Two Wedges that Crimp RNA. *Mol. Cell* (2009) doi:10.1016/j.molcel.2009.07.032.

73. Tijerina, P., Bhaskaran, H. & Russell, R. Nonspecific binding to structured RNA and preferential unwinding of an exposed helix by the CYT-19 protein, a DEAD-box RNA chaperone. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 16698–16703 (2006).

74. Collins, R. *et al.* The DEXD/H-box RNA helicase DDX19 is regulated by an {alpha}-helical switch. *J. Biol. Chem.* **284**, 10296–300 (2009).

75. Von Moeller, H., Basquin, C. & Conti, E. The mRNA export protein DBP5 binds RNA and the cytoplasmic nucleoporin NUP214 in a mutually exclusive manner. *Nat. Struct. Mol. Biol.* (2009) doi:10.1038/nsmb.1561.

76. Dufau, M. L. & Tsai-Morris, C. H. Gonadotropin-regulated testicular helicase (GRTH/DDX25): an essential regulator of spermatogenesis. *Trends in Endocrinology and Metabolism* (2007) doi:10.1016/j.tem.2007.09.001.

77. Fleckner, J., Zhang, M., Valcárcel, J. & Green, M. R. U2AF65 recruits a novel human DEAD box protein required for the U2 snRNP-branchpoint interaction. *Genes Dev.* (1997)

doi:10.1101/gad.11.14.1864.

78. Shen, H. UAP56- a key player with surprisingly diverse roles in pre-mRNA splicing and nuclear export. *BMB Reports* (2009) doi:10.5483/BMBRep.2009.42.4.185.

79. Shen, H. *et al.* Distinct activities of the DExD/H-box splicing factor hUAP56 facilitate stepwise assembly of the spliceosome. *Genes Dev.* (2008) doi:10.1101/gad.1657308.

80. Iwasaki, Y. W., Siomi, M. C. & Siomi, H. PIWI-Interacting RNA: Its Biogenesis and Functions. *Annu. Rev. Biochem.* (2015) doi:10.1146/annurev-biochem-060614-034258.

81. Zhang, G. *et al.* Co-dependent Assembly of Drosophila piRNA Precursor Complexes and piRNA Cluster Heterochromatin. *Cell Rep.* (2018) doi:10.1016/j.celrep.2018.08.081.

82. Zhang, F. *et al.* UAP56 couples piRNA clusters to the perinuclear transposon silencing machinery. *Cell* (2012) doi:10.1016/j.cell.2012.09.040.

83. ElMaghraby, M. F. *et al.* A Heterochromatin-Specific RNA Export Pathway Facilitates piRNA Production. *Cell* (2019) doi:10.1016/j.cell.2019.07.007.

84. Smith, D. J. *et al.* Mapping the antigenic and genetic evolution of influenza virus. *Science (80-.).* (2004) doi:10.1126/science.1097211.

85. Webster, R. G. 1918 Spanish influenza: The secrets remain elusive. *Proceedings of the National Academy of Sciences of the United States of America* (1999) doi:10.1073/pnas.96.4.1164.

86. Konishi, T. Re-evaluation of the evolution of influenza H1 viruses using direct PCA. *Sci. Rep.* (2019) doi:10.1038/s41598-019-55254-z.

87. A revision of the system of nomenclature for influenza viruses: a WHO memorandum. *Bull. World Health Organ.* **58**, 585–591 (1980).

88. Webster, R. G., Bean, W. J., Gorman, O. T., Chambers, T. M. & Kawaoka, Y. Evolution and ecology of influenza A viruses. *Microbiological Reviews* vol. 56 152–179 (1992).

89. Berkaloff, A. Etude au microscope Electronique de la particule virale de myxovirus influenza B (souche Lee) comparison des resultats en "coloration negative" et en coupe. *J. Microsc. (Paris).* **2**, 583–604 (1963).

90. Pons, M. W., Schulze, I. T. & Hirst, G. K. Isolation and characterization of the ribonucleoprotein of influenza virus. *Virology* **39**, 250–259 (1969).

91. Schulze, I. T. The structure of influenza virus. II. A model based on the morphology and composition of subviral particles. *Virology* **47**, 181–196 (1972).

92. Schulze, I. T. The structure of influenza virus. I. The polypeptides of the virion. *Virology* **42**, 890–904 (1970).

93. Dou, D., Revol, R., Östbye, H., Wang, H. & Daniels, R. Influenza A virus cell entry, replication, virion assembly and movement. *Frontiers in Immunology* (2018) doi:10.3389/fimmu.2018.01581.

94. Zhang, K. *et al.* Dissection of Influenza A Virus M1 Protein: pH-Dependent Oligomerization of N-Terminal Domain and Dimerization of C-Terminal Domain. doi:10.1371/journal.pone.0037786.

95. Bui, M., Whittaker, G. & Helenius, A. Effect of M1 protein and low pH on nuclear transport of influenza virus ribonucleoproteins. *J. Virol.* (1996)

doi:10.1128/jvi.70.12.8391-8401.1996.

96. Kielian, M. & Rey, F. A. Virus membrane-fusion proteins: More than one way to make a hairpin. *Nature Reviews Microbiology* (2006) doi:10.1038/nrmicro1326.

97. Thomaston, J. L. *et al.* X-ray Crystal Structures of the Influenza M2 Proton Channel Drug-Resistant V27A Mutant Bound to a Spiro-Adamantyl Amine Inhibitor Reveal the Mechanism of Adamantane Resistance. *Biochemistry* (2020) doi:10.1021/acs.biochem.9b00971.

98. O'Neill, R. E., Jaskunas, R., Blobel, G., Palese, P. & Moroianu, J. Nuclear import of influenza virus RNA can be mediated by viral nucleoprotein and transport factors required for protein import. *J. Biol. Chem.* (1995) doi:10.1074/jbc.270.39.22701.

99. Kemler, I., Whittaker, G. & Helenius, A. Nuclear Import of Microinjected Influenza Virus Ribonucleoproteins. *Virology* (1994) doi:10.1006/viro.1994.1432.

100. Wang, P. *et al.* The NPI-1/NPI-3 (karyopherin alpha) binding site on the influenza a virus nucleoprotein NP is a nonconventional nuclear localization signal. *J. Virol.* **71**, 1850–6 (1997).

101. Pflug, A., Guilligay, D., Reich, S. & Cusack, S. Structure of influenza A polymerase bound to the viral RNA promoter. *Nature* **516**, 355–360 (2014).

102. Ulmanen, I., Broni, B. A. & Krug, R. M. Role of two of the influenza virus core P proteins in recognizing cap 1 structures (m7GpppNm) on RNAs and in initiating viral RNA transcription (UV-induced crosslinkdng/two-dimensional gel electrophoresis). *Biochemistry* (1981).

103. Bouloy, M., Plotch, S. J. & Krug, R. M. Globin mRNAs are primers for the transcription of influenza viral RNA in vitro. *Proc. Natl. Acad. Sci. U. S. A.* (1978) doi:10.1073/pnas.75.10.4886.

104. Plotch, S. J., Bouloy, M. & Krug, R. M. Transfer of 5'-terminal cap of globin mRNA to influenza viral complementary RNA during transcription in vitro. *Proc. Natl. Acad. Sci. U. S. A.* (1979) doi:10.1073/pnas.76.4.1618.

105. Beaton, A. R. & Krug, R. M. Selected host cell capped RNA fragments prime influenza viral RNA transcription in vivo. *Nucleic Acids Res.* (1981) doi:10.1093/nar/9.17.4423.

106. Plotch, S. J., Bouloy, M., Ulmanen, I. & Krug, R. M. A unique cap(m7GpppXm)-dependent influenza virion endonuclease cleaves capped RNAs to generate the primers that initiate viral RNA transcription. *Cell* **23**, 847–858 (1981).

107. Blass, D., Patzelt, E. & Kuechler, E. Identification of the cap binding protein of influenza virus. *Nucleic Acids Res.* (1982) doi:10.1093/nar/10.15.4803.

108. Guilligay, D. *et al.* The structural basis for cap binding by influenza virus polymerase subunit PB2. *Nat. Struct. Mol. Biol.* **15**, 500–506 (2008).

109. Lukarska, M. *et al.* Structural basis of an essential interaction between influenza polymerase and Pol II CTD. *Nature* (2017) doi:10.1038/nature20594.

110. Li, M. L., Rao, P. & Krug, R. M. The active sites of the influenza cap-dependent endonuclease are on different polymerase subunits. *EMBO J.* (2001) doi:10.1093/emboj/20.8.2078.

111. Dias, A. *et al.* The cap-snatching endonuclease of influenza virus polymerase resides in

the PA subunit. *Nature* (2009) doi:10.1038/nature07745.

112. York, A. & Fodor, E. Biogenesis, assembly and export of viral messenger ribonucleoproteins in the influenza A virus infected cell. *RNA Biol.* (2013) doi:10.4161/rna.25356.

113. Mor, A. *et al.* Influenza virus mRNA trafficking through host nuclear speckles. *Nat. Microbiol.* (2016) doi:10.1038/nmicrobiol.2016.69.

114. Zhang, K. *et al.* Structural basis for influenza virus NS1 protein block of mRNA nuclear export. *Nature Microbiology* (2019) doi:10.1038/s41564-019-0482-x.

115. Gerl, M. J. *et al.* Quantitative analysis of the lipidomes of the influenza virus envelope and MDCK cell apical membrane. *J. Cell Biol.* (2012) doi:10.1083/jcb.201108175.

116. Eisfeld, A. J., Neumann, G. & Kawaoka, Y. At the centre: influenza A virus ribonucleoproteins. *Nat. Rev. Microbiol.* **13**, 28–41 (2015).

117. Poole, E., Elton, D., Medcalf, L. & Digard, P. Functional domains of the influenza A virus PB2 protein: Identification of NP- and PB1-binding sites. *Virology* **321**, 120–133 (2004).

118. Robb, N. C., Vreede, F. T., Smith, M. & Fodor, E. NS2/NEP protein regulates transcription and replication of the influenza virus RNA genome. *J. Gen. Virol.* **90**, 1398–1407 (2009).

119. Reich, S. *et al.* Structural insight into cap-snatching and RNA synthesis by influenza polymerase. *Nature* **516**, 361–366 (2014).

120. Butcher, S. J., Grimes, J. M., Makeyev, E. V., Bamford, D. H. & Stuart, D. I. A mechanism for initiating RNA-dependent RNA polymerization. *Nature* **410**, 235–240

(2001).

121. Caillet-Saguy, C., Lim, S. P., Shi, P. Y., Lescar, J. & Bressanelli, S. Polymerases of hepatitis C viruses and flaviviruses: Structural and mechanistic insights and drug development. *Antiviral Research* (2014) doi:10.1016/j.antiviral.2014.02.006.

122. Deng, T., Vreede, F. T. & Brownlee, G. G. Different De Novo Initiation Strategies Are Used by Influenza Virus RNA Polymerase on Its cRNA and Viral RNA Promoters during Viral RNA Replication. *J. Virol.* (2006) doi:10.1128/jvi.80.5.2337-2348.2006.

123. Vreede, F. T., Gifford, H. & Brownlee, G. G. Role of Initiating Nucleoside Triphosphate Concentrations in the Regulation of Influenza Virus Replication and Transcription. *J. Virol.* (2008) doi:10.1128/jvi.00627-08.

124. Fodor, E. The RNA polymerase of influenza A virus: Mechanisms of viral transcription and replication. *Acta Virol.* (2013) doi:10.4149/av_2013_02_113.

125. Jorba, N., Coloma, R. & Ortín, J. Genetic trans-complementation establishes a new model for influenza virus RNA transcription and replication. *PLoS Pathog.* (2009) doi:10.1371/journal.ppat.1000462.

126. Ye, Z., Liu, T., Offringa, D. P., McInnis, J. & Levandowski, R. A. Association of Influenza Virus Matrix Protein with Ribonucleoproteins. *J. Virol.* **73**, 7467–7473 (1999).

127. O'Neill, R. E., Talon, J. & Palese, P. The influenza virus NEP (NS2 protein) mediates the nuclear export of viral ribonucleoproteins. *EMBO J.* (1998) doi:10.1093/emboj/17.1.288.

128. Huang, S. *et al.* A Second CRM1-Dependent Nuclear Export Signal in the Influenza A Virus NS2 Protein Contributes to the Nuclear Export of Viral Ribonucleoproteins. *J.*

Virol. (2013) doi:10.1128/jvi.06519-11.

129. Mathew, C. & Ghildyal, R. CRM1 inhibitors for antiviral therapy. *Frontiers in Microbiology* (2017) doi:10.3389/fmicb.2017.01171.

130. Momose, F. *et al.* Apical transport of influenza A virus ribonucleoprotein requires Rab11-positive recycling endosome. *PLoS One* (2011) doi:10.1371/journal.pone.0021123.

131. Amorim, M. J. A comprehensive review on the interaction between the host GTPase Rab11 and influenza A virus. *Frontiers in Cell and Developmental Biology* (2019) doi:10.3389/fcell.2018.00176.

132. Zhang, J., Pekosz, A. & Lamb, R. A. Influenza Virus Assembly and Lipid Raft Microdomains: a Role for the Cytoplasmic Tails of the Spike Glycoproteins. *J. Virol.* (2000) doi:10.1128/jvi.74.10.4634-4644.2000.

133. Rossman, J. S. & Lamb, R. A. Influenza virus assembly and budding. *Virology* (2011) doi:10.1016/j.virol.2010.12.003.

134. Rossman, J. S. *et al.* Influenza Virus M2 Ion Channel Protein Is Necessary for Filamentous Virion Formation. *J. Virol.* (2010) doi:10.1128/jvi.00119-10.

135. Kim, S. S. *et al.* Cholesterol-Dependent Conformational Exchange of the C-Terminal Domain of the Influenza A M2 Protein. *Biochemistry* **54**, 7157–7167 (2015).

136. Enkirch, T. *et al.* Identification and in vivoefficacy assessment of approved orally bioavailable human host protein-targeting drugs with broad anti-influenza a activity. *Front. Immunol.* (2019) doi:10.3389/fimmu.2019.01097.

137. Satterly, N. *et al.* Influenza virus targets the mRNA export machinery and the nuclear pore

complex. *Proc. Natl. Acad. Sci. U. S. A.* (2007) doi:10.1073/pnas.0610977104.

138. Das, K. *et al.* Structural basis for suppression of a host antiviral response by influenza a virus. *Proc. Natl. Acad. Sci. U. S. A.* (2008) doi:10.1073/pnas.0805213105.

139. Chen, Z., Li, Y. & Krug, R. M. Influenza A virus NS1 protein targets poly(A)-binding protein II of the cellular 3'-end processing machinery. *EMBO J.* **18**, 2273–83 (1999).

140. Takeuchi, O. & Akira, S. Pattern Recognition Receptors and Inflammation. *Cell* (2010) doi:10.1016/j.cell.2010.01.022.

141. Clemens, M. J. & Williams, B. R. G. Inhibition of cell-free protein synthesis by pppA2′ p5′ A2′ p5′ A: a novel oligonucleotide synthesized by interferon-treated L cell extracts. *Cell* (1978) doi:10.1016/0092-8674(78)90329-X.

142. Cheng, A., Wong, S. M. & Yuan, Y. A. Structural basis for dsRNA recognition by NS1 protein of influenza A virus. *Cell Res.* (2009) doi:10.1038/cr.2008.288.

143. Gack, M. U. *et al.* Influenza A Virus NS1 Targets the Ubiquitin Ligase TRIM25 to Evade Recognition by the Host Viral RNA Sensor RIG-I. *Cell Host Microbe* (2009) doi:10.1016/j.chom.2009.04.006.

144. Mibayashi, M. *et al.* Inhibition of Retinoic Acid-Inducible Gene I-Mediated Induction of Beta Interferon by the NS1 Protein of Influenza A Virus. *J. Virol.* (2007) doi:10.1128/jvi.01265-06.

145. Li, S., Min, J. Y., Krug, R. M. & Sen, G. C. Binding of the influenza A virus NS1 protein to PKR mediates the inhibition of its activation by either PACT or double-stranded RNA. *Virology* (2006) doi:10.1016/j.virol.2006.01.005.

146. Andy Ka-Leung, N., Jia-Huai, W. & Pang-Chui, S. Structure and sequence analysis of influenza A virus nucleoprotein. (2009) doi:10.1007/s11427-009-0064-x.

147. Yamanaka, K., Ishihama, A. & Nagata, K. Reconstitution of influenza virus RNA-nucleoprotein complexes structurally resembling native viral ribonucleoprotein cores. *J. Biol. Chem.* (1990).

148. Baudin, F., Bach, C., Cusack, S. & Ruigrok, R. W. Structure of influenza virus RNP. I. Influenza virus nucleoprotein melts secondary structure in panhandle RNA and exposes the bases to the solvent. *EMBO J.* (1994) doi:10.1002/j.1460-2075.1994.tb06614.x.

149. Zheng, W. & Tao, Y. J. Structure and assembly of the influenza A virus ribonucleoprotein complex. *FEBS Letters* (2013) doi:10.1016/j.febslet.2013.02.048.

150. Hsu, M. T., Parvin, J. D., Gupta, S., Krystal, M. & Palese, P. Genomic RNAs of influenza viruses are held in a circular conformation in virions and in infected cells by a terminal panhandle. *Proc. Natl. Acad. Sci. U. S. A.* (1987) doi:10.1073/pnas.84.22.8140.

151. Klumpp, K. Roles of the influenza virus polymerase and nucleoprotein in forming a functional RNP structure. *EMBO J.* **16**, 1248–1257 (1997).

152. Hagen, M., Chung, T. D., Butcher, J. A. & Krystal, M. Recombinant influenza virus polymerase: requirement of both 5’ and 3’ viral ends for endonuclease activity. *J. Virol.* (1994) doi:10.1128/jvi.68.3.1509-1515.1994.

153. Ye, Q., Krug, R. M. & Tao, Y. J. The mechanism by which influenza A virus nucleoprotein forms oligomers and binds RNA. *Nature* **444**, 1078–1082 (2006).

154. Chenavas, S. *et al.* Monomeric Nucleoprotein of Influenza A Virus. *PLoS Pathog.* **9**,

e1003275 (2013).

155. Tarus, B. *et al.* Oligomerization paths of the nucleoprotein of influenza A virus. *Biochimie* **94**, 776–785 (2012).

156. Krug, R. M., Alonso-Caplen, F. V., Julkunen, I. & Katze, M. G. Expression and Replication of the Influenza Virus Genome. in *The Influenza Viruses* 89–152 (Springer US, 1989). doi:10.1007/978-1-4613-0811-9_2.

157. Privalsky, M. L. & Penhoet, E. E. Influenza virus proteins: Identity, synthesis, and modification analyzed by two dimensional gel electrophoresis. *Proc. Natl. Acad. Sci. U. S. A.* (1978) doi:10.1073/pnas.75.8.3625.

158. Arrese, M. & Portela, A. Serine 3 is critical for phosphorylation at the N-terminal end of the nucleoprotein of influenza virus A/Victoria/3/75. *J. Virol.* (1996) doi:10.1128/jvi.70.6.3385-3391.1996.

159. Cui, L. *et al.* Phosphorylation Status of Tyrosine 78 Residue Regulates the Nuclear Export and Ubiquitination of Influenza A Virus Nucleoprotein. *Front. Microbiol.* (2019) doi:10.3389/fmicb.2019.01816.

160. Li, Y. *et al.* Phosphorylation and dephosphorylation of threonine 188 in nucleoprotein is crucial for the replication of influenza A virus. *Virology* **520**, 30–38 (2018).

161. Turrell, L., Hutchinson, E. C., Vreede, F. T. & Fodor, E. Regulation of Influenza A Virus Nucleoprotein Oligomerization by Phosphorylation Downloaded from. *J. Virol. Febr.* **89**, 2020 (2015).

162. Hutchinson, E. C. *et al.* Mapping the Phosphoproteome of Influenza A and B Viruses by

Mass Spectrometry. *PLoS Pathog.* **8**, e1002993 (2012).

163. Kistner, O., Muller, K. & Scholtissek, C. Differential Phosphorylation of the Nucleoprotein of Influenza A Viruses. *J. Gen. Virol.* **70**, 2421–2431 (1989).

164. Biswas, S. K., Boutz, P. L. & Nayak, D. P. Influenza Virus Nucleoprotein Interacts with Influenza Virus Polymerase Proteins. *J. Virol.* **72**, 5493–5501 (1998).

165. Marklund, J. K., Ye, Q., Dong, J., Tao, Y. J. & Krug, R. M. Sequence in the Influenza A Virus Nucleoprotein Required for Viral Polymerase Binding and RNA Synthesis. (2012) doi:10.1128/JVI.00014-12.

166. Ng, A. K. L. *et al.* Influenza polymerase activity correlates with the strength of interaction between nucleoprotein and PB2 through the host-specific residue K/E627. *PLoS One* (2012) doi:10.1371/journal.pone.0036415.

167. Noda, T. *et al.* Three-dimensional analysis of ribonucleoprotein complexes in influenza A virus. *Nat. Commun.* (2012) doi:10.1038/ncomms1647.

168. Williams, G. D. *et al.* Nucleotide resolution mapping of influenza A virus nucleoprotein-RNA interactions reveals RNA features required for replication. *Nat. Commun.* (2018) doi:10.1038/s41467-018-02886-w.

169. Lee, N. *et al.* Genome-wide analysis of influenza viral RNA and nucleoprotein association. *Nucleic Acids Res.* (2017) doi:10.1093/nar/gkx584.

170. Hutchinson, E. C., von Kirchbach, J. C., Gog, J. R. & Digard, P. Genome packaging in influenza A virus. *Journal of General Virology* (2010) doi:10.1099/vir.0.017608-0.

171. Moreira, É. A. *et al.* A conserved influenza A virus nucleoprotein code controls specific

viral genome packaging. *Nat. Commun.* (2016) doi:10.1038/ncomms12861.

172. Bolte, H., Rosu, M. E., Hagelauer, E., García-Sastre, A. & Schwemmle, M. Packaging of the Influenza Virus Genome Is Governed by a Plastic Network of RNA- and Nucleoprotein-Mediated Interactions. *J. Virol.* **93**, (2018).

173. Momose, F., Handa, H. & Nagata, K. Identification of host factors that regulate the influenza virus RNA polymerase activity. *Biochimie* **78**, 1103–8 (1996).

174. Minakuchi, M. *et al.* Pre-mRNA Processing Factor Prp18 Is a Stimulatory Factor of Influenza Virus RNA Synthesis and Possesses Nucleoprotein Chaperone Activity. *J. Virol.* (2017) doi:10.1128/jvi.01398-16.

175. Naito, T. *et al.* An influenza virus replicon system in yeast identified Tat-SF1 as a stimulatory host factor for viral RNA synthesis. *Proc. Natl. Acad. Sci. U. S. A.* (2007) doi:10.1073/pnas.0705856104.

176. Hulver, M. J. *et al.* Human Tat-specific factor 1 binds the HIV-1 genome and selectively transports HIV-1 RNAs. *Mol. Biol. Rep.* **47**, 1759–1772 (2020).

177. Cortez, D., Glick, G. & Elledge, S. J. Minichromosome maintenance proteins are direct targets of the ATM and ATR checkpoint kinases. *Proc. Natl. Acad. Sci. U. S. A.* (2004) doi:10.1073/pnas.0403410101.

178. Kawaguchi, A. & Nagata, K. De novo replication of the influenza virus RNA genome is regulated by DNA replicative helicase, MCM. *EMBO J.* (2007) doi:10.1038/sj.emboj.7601881.

179. Kawaguchi, A., Momose, F. & Nagata, K. Replication-Coupled and Host Factor-Mediated

Encapsidation of the Influenza Virus Genome by Viral Nucleoprotein. *J. Virol.* (2011) doi:10.1128/JVI.00277-11.

180. Fong, Y. W. & Zhou, Q. Stimulatory effect of splicing factors on transcriptional elongation. *Nature* (2001) doi:10.1038/414929a.

181. Yankulov, K. *et al.* MCM Proteins Are Associated with RNA Polymerase II Holoenzyme. *Mol. Cell. Biol.* (1999) doi:10.1128/mcb.19.9.6154.

182. Holland, L. *et al.* Distinct parts of minichromosome maintenance protein 2 associate with histone H3/H4 and RNA polymerase II holoenzyme. *Eur. J. Biochem.* (2002) doi:10.1046/j.1432-1033.2002.03224.x.

183. Gauthier, L. *et al.* The role of the carboxyterminal domain of RNA polymerase II in regulating origins of DNA replication in Saccharomyces cerevisiae. *Genetics* (2002).

184. Landeras-Bueno, S., Jorba, N., Pérez-Cidoncha, M. & Ortín, J. The splicing factor proline-glutamine rich (SFPQ/PSF) is involved in influenza virus transcription. *PLoS Pathog.* (2011) doi:10.1371/journal.ppat.1002397.

185. Jorba, N. *et al.* Analysis of the interaction of influenza virus polymerase complex with human cell factors. *Proteomics* (2008) doi:10.1002/pmic.200700508.

186. Beaton, A. R. & Krug, R. M. Transcription antitermination during influenza viral template RNA synthesis requires the nucleocapsid protein and the absence of a 5' capped end. *Proc. Natl. Acad. Sci.* **83**, 6282–6286 (1986).

187. Momose, F. *et al.* Cellular splicing factor RAF-2p48/NPI-5/BAT1/UAP56 interacts with the influenza virus nucleoprotein and enhances viral RNA synthesis. *J. Virol.* **75**, 1899–

908 (2001).

188. O'Neill, R. E. & Palese, P. NPI-1, the human homolog of SRP-1, Interacts with influenza virus nucleoprotein. *Virology* **206**, 116–125 (1995).

189. Palese, P., Wang, P., Wolff, T. & O'Neill, R. E. Host-Viral Protein-Protein Interactions in Influenza Virus Replication. in *Molecular aspects of host-pathogen interactions* (eds. McCrae, M. A., Saunders, J. R., Smyth, C. J. & Snow, N. D.) 327–340 (Cambridge University Press, 1997).

190. Hu, Y., Gor, V., Morikawa, K., Nagata, K. & Kawaguchi, A. Cellular splicing factor UAP56 stimulates trimeric NP formation for assembly of functional influenza viral ribonucleoprotein complexes. *Sci. Rep.* **7**, 14053 (2017).

191. Moeller, A., Kirchdoerfer, R. N., Potter, C. S., Carragher, B. & Wilson, I. A. Organization of the Influenza Virus Replication Machinery. *Science (80-.).* **338**, 1631–1634 (2012).

192. Ye, Q. *et al.* Biochemical and structural evidence in support of a coherent model for the formation of the double-helical influenza A virus ribonucleoprotein. *MBio* **4**, e00467-12 (2012).

193. Arai, Y. *et al.* Novel Polymerase Gene Mutations for Human Adaptation in Clinical Isolates of Avian H5N1 Influenza Viruses. *PLoS Pathog.* **12**, e1005583 (2016).

194. Arai, Y. *et al.* Multiple polymerase gene mutations for human adaptation occurring in Asian H5N1 influenza virus clinical isolates. *Sci. Rep.* **8**, 1–13 (2018).

195. Manz, B., Schwemmle, M. & Brunotte, L. Adaptation of Avian Influenza A Virus Polymerase in Mammals To Overcome the Host Species Barrier. *J. Virol.* **87**, 7200–7209

(2013).

196. H, L. *et al.* Molecular Cell Biology. in *Molecular Cell Biology* (W. H. Freeman and Company, 2000).

197. Scheuermann, T. H., Padrick, S. B., Gardner, K. H. & Brautigam, C. A. On the acquisition and analysis of microscale thermophoresis data. *Anal. Biochem.* **496**, 79–93 (2016).

198. Asmari, M., Ratih, R., Alhazmi, H. A. & El Deeb, S. Thermophoresis for characterizing biomolecular interaction. *Methods* **146**, 107–119 (2018).

199. Gradinaru, C. C., Marushchak, D. O., Samim, M. & Krull, U. J. Fluorescence anisotropy: From single molecules to live cells. *Analyst* (2010) doi:10.1039/b920242k.

200. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* (1997) doi:10.1016/S0076-6879(97)76066-X.

201. Liebschner, D. *et al.* Macromolecular structure determination using X-rays, neutrons and electrons: Recent developments in Phenix. *Acta Crystallogr. Sect. D Struct. Biol.* (2019) doi:10.1107/S2059798319011471.

202. Ma, Z. *et al.* Supporting tool suite for production proteomics. *Bioinformatics* **27**, 3214–3215 (2011).

203. Götze, M. *et al.* StavroX-A software for analyzing crosslinked products in protein interaction studies. *J. Am. Soc. Mass Spectrom.* (2012) doi:10.1007/s13361-011-0261-2.

204. Dasari, S. *et al.* TagRecon: High-throughput mutation identification through sequence tagging. *J. Proteome Res.* (2010) doi:10.1021/pr900850m.

205. Schneider, C. A., Rasband, W. S. & Eliceiri, K. W. NIH Image to ImageJ: 25 years of image analysis. *Nature Methods* (2012) doi:10.1038/nmeth.2089.

206. Hanke, L. *et al.* The antiviral mechanism of an influenza a virus nucleoprotein-specific single-domain antibody fragment. *MBio* (2016) doi:10.1128/mBio.01569-16.

207. Gerritz, S. W. *et al.* Inhibition of influenza virus replication via small molecules that induce the formation of higher-order nucleoprotein oligomers. *Proc. Natl. Acad. Sci. U. S. A.* (2011) doi:10.1073/pnas.1107906108.

208. Pang, B. *et al.* Structural Characterization of H1N1 Nucleoprotein-Nucleozin Binding Sites. *Sci. Rep.* **6**, 29684 (2016).

209. Li, Z. *et al.* Mutational Analysis of Conserved Amino Acids in the Influenza A Virus Nucleoprotein. *J. Virol.* **83**, 4153–4162 (2009).

210. Wisskirchen, C., Ludersdorfer, T. H., Muller, D. A., Moritz, E. & Pavlovic, J. The Cellular RNA Helicase UAP56 Is Required for Prevention of Double-Stranded RNA Formation during Influenza A Virus Infection. *J. Virol.* **85**, 8646–8655 (2011).

211. Shen, Y.-F. *et al.* E339...R416 salt bridge of nucleoprotein as a feasible target for influenza virus inhibitors. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 16515–20 (2011).

212. Hung, S. J. *et al.* Genetic variations on 31 and 450 residues of influenza A nucleoprotein affect viral replication and translation. *J. Biomed. Sci.* (2020) doi:10.1186/s12929-019-0612-z.

213. Serna Martin, I. *et al.* A Mechanism for the Activation of the Influenza Virus Transcriptase. *Mol. Cell* (2018) doi:10.1016/j.molcel.2018.05.011.

214. Resa-Infante, P., Jorba, N., Coloma, R. & Ortin, J. The influenza virus RNA synthesis machine. *RNA Biol.* **8**, 207–215 (2011).